AD 680785

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

TM-738/049/00

Distribution Of This Document Is Unlimited
It May Be Released To The Clearinghouse,
Department Of Commerce, For Sale To The General Public

Substitution in Families of Languages

Seymour Ginsburg Edwin H. Spanier

SCIENTIFIC REPORT NO. 22

16 September 1968

DDC
RECEIVED
JAN 22 1969
C

Prepared for
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH, USAF
BEDFORD, MASSACHUSETTS
and
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH, USAF
ARLINGTON, VIRGINIA

Contract Monitor Thomas V. Griffiths
Data Sciences Laboratory
Monitored by R. Swanson, SRIR (AFOSR)

Supported by
SDC's Independent Research Program.

Contract F1962867C0008, Programming (Algorithmic) Languages, Project No. 5632,
Task No. 563205, and Grant No. AF AFOSR 1203-67
Work Unit No. 56320501

41

CFSTI
DDC
AVAILABILITY CODES
DIST. AVAIL. and/or SPECIAL

Qualified requestors may obtain additional copies from the Defense Documentation Center. All others should apply to the Clearinghouse for Federal Scientific and Technical Information.

AFCRL-68-0592

TM-738/049/00

SDC TM SERIES

The work reported herein was supported by SDC and Contract F1962867C0008, Programming (Algorithmic) Languages; Project No. 5632, Task No. 563205; Work Unit No. 56320501; and Grant No. AF-AFOSR-1203-67

SCIENTIFIC REPORT NO. 22

Substitution in Families of Languages

by

Seymour Ginsburg*
and
Edwin H. Spanier**

16 September 1968

Monitored by: Contract Monitor
Thomas V. Griffiths
Data Sciences Laboratory
R. Swanson, SRIR (AFOSR)

Prepared for
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH, USAF
BEDFORD MASSACHUSETTS 01730
and
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH, USAF
ARLINGTON, VIRGINIA

SYSTEM
DEVELOPMENT
CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

*System Development Corporation and University of Southern California

**University of California at Berkeley and consultant, System Development Corporation

ABSTRACT

The effect of substitution in families of languages, especially AFL, is considered. Among the main results shown are the following: The substitution of one AFL into another is an AFL. Under suitable hypotheses, the AFL generated by the family obtained from the substitution of one family into another, is the family obtained from the substitution of the corresponding AFL. A condition is given for the AFL generated by the substitution closure of a family to be the substitution closure of the AFL generated by the family.

BLANK PAGE

SUBSTITUTION IN FAMILIES OF LANGUAGES *

INTRODUCTION

In an earlier paper [6], the authors were led to consider the family of languages obtained from the family of linear context-free languages by iterated substitution. This in turn suggested to us a study of the substitution of one arbitrary family of languages into another and is the subject of the present work.

Recently the notion of an AFL (abstract family of languages) was introduced [3] as an abstraction of many of the formal languages of concern to computer science. In particular, it was shown in [3] that there is an intimate connection between AFL and the families of languages accepted by families of one-way nondeterministic acceptors. Thus AFL play a special role among arbitrary families of languages, at least for device theory. The theorems contained herein are concerned with the relationship of AFL and substitution. These theorems, in turn, are based on lesser results which are concerned with the relation of substitution to other operations. These lesser results are formulated in terms of arbitrary families of languages because (1) many of them are interesting in their own right and may have other applications, and (2) they isolate the difficulty inherent in the proofs of the main results.

*Research sponsored in part by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, USAF, under contract F1962867C0008, and by the Air Force Office of Scientific Research, Office of Aerospace Research, USAF, under AFOSR Grant No. AF-AFOSR-1203-67.

The paper is organized into four sections. The first one is devoted to general concepts of families of languages, to a proof that under a mild condition substitution satisfies a kind of associativity, and to a formulation of the AFL properties in terms of substitution. Section two is concerned with showing that the substitution of one AFL into another is an AFL. Section three contains a proof that under suitable hypotheses the AFL generated by the family obtained from the substitution of one family into another is the family obtained from the substitution of the corresponding AFL.

Section four considers iterated substitution and the substitution closure of one family with respect to another. A special case is the substitution closure of a family. It is shown that the various substitution closures give AFL when applied to AFL, and a condition is given for the AFL generated by the substitution closure of a family to be the substitution closure of the AFL generated by the family. A condition is also given for the substitution closure of a family, not necessarily an AFL, to be an AFL. This specializes to the case of the family of linear context-free languages and implies that the substitution closure of this family is a full AFL (a result obtained by other means in [6]).

Section 1. Families of languages

In this section we review some concepts about families of languages. We also examine certain methods of constructing new families from old, especially substituting one family into another. Finally, we consider the concept of an abstract family of languages and reformulate it in terms of substitution.

<u>Definition</u>. A <u>family of languages</u> is a pair $(\Sigma, \mathcal{L})$, or $\mathcal{L}$ when Σ is understood, where

(1) Σ is an infinite set of symbols,

(2) for each L in $\mathcal{L}$ there is a finite set $\Sigma_1 \subseteq \Sigma$ such that[(1)] $L \subseteq \Sigma_1^*$,

and (3) $L \neq \emptyset$ for some L in $\mathcal{L}$.

<u>Notation</u>. Given Σ in $\mathcal{L}$, Σ_L will denote the smallest set Σ_1 such that $L \subseteq \Sigma_1^*$.

Henceforth, Σ will always denote a given infinite set of symbols and Σ with a subscript a finite subset of Σ. All symbols given or constructed will be assumed in Σ.

We now distinguish some elementary conditions for families of languages.

<u>Definition</u>. A family of languages $\mathcal{L}$ is said to be

(1) <u>symmetric</u> if it is invariant under all permutations of Σ.

(2) <u>ϵ-free</u> if each L in $\mathcal{L}$ is ϵ-free (i.e., ϵ is not in L).

(3) <u>nontrivial</u> if there is some L in $\mathcal{L}$ containing a non-ϵ word.

We shall be interested in various operations on languages and families of languages. We first present two operations on pairs of families.

<u>Notation</u>. Given families of languages $\mathcal{L}_1$ and $\mathcal{L}_2$, let

(1) $\mathcal{L}_1 \wedge \mathcal{L}_2 = \{L_1 \cap L_2 / L_1 \text{ in } \mathcal{L}_1, L_2 \text{ in } \mathcal{L}_2\}$.

(2) Sub $(\mathcal{L}_1, \mathcal{L}_2)$ be the family obtained by substituting languages of $\mathcal{L}_1$ into languages of $\mathcal{L}_2$, i.e., the family of all sets $\tau(L_2)$, where L_2 is in $\mathcal{L}_2$ and τ is a substitution[(2)] such that $\tau(a)$ is in $\mathcal{L}_1$ for each a in Σ_{L_2}.

---

(1) Σ_1^* is the free semigroup with identity ϵ generated by Σ_1, i.e., the set of all finite strings $a_1 \dots a_n$, each a_i in Σ_1. Each element of Σ_1^* is called a <u>word</u> of Σ_1^*.

(2) Let $L \subseteq \Sigma_3^*$ and for each a in Σ_3 let $L_a \subseteq \Sigma_a^*$. Let τ be the function defined on Σ_3^* by $\tau(\epsilon) = \{\epsilon\}$, $\tau(a) = L_a$ for each a in Σ_3, and $\tau(a_1 \dots a_n) = \tau(a_1) \dots \tau(a_n)$ for each a_i in Σ_3 and $k \geq 1$. Then τ is called a <u>substitution</u>. τ is extended 2^{Σ_3} by defining $\tau(X) = \bigcup_{x \text{ in } X} \tau(x)$ for all $X \subseteq \Sigma_3^*$.

As is evident from the title, our interest here is in substitutions in families of languages.

We next present two operations on a family of languages.

<u>Notation</u>. For each family $\mathcal{L}$ let

(1) $\mathrm{Hom}(\mathcal{L}) = \{h(L)/L \text{ in } \mathcal{L},\ h \text{ a homomorphism}^{(3)} \text{ on } L\}$, and

(2) $\mathrm{Hom}_r(\mathcal{L}) = \{h(L)/L \text{ in } \mathcal{L},\ h \text{ a restricted homomorphism on } L^{(4)}\}$.

Clearly $\mathrm{Hom}(\mathcal{L})$ and $\mathrm{Hom}_r(\mathcal{L})$ are monotonically increasing in $\mathcal{L}^{(5)}$, and both $\mathcal{L}_1 \wedge \mathcal{L}_2$ and $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ are monotonically increasing in each of the families $\mathcal{L}_1$ and $\mathcal{L}_2$. For each family $\mathcal{L}$, $\mathrm{Hom}(\mathcal{L})$ and $\mathrm{Hom}_r(\mathcal{L})$ are symmetric, $\mathrm{Hom}\ \mathrm{Hom}(\mathcal{L}) = \mathrm{Hom}(\mathcal{L})$, and $\mathrm{Hom}_r\ \mathrm{Hom}_r(\mathcal{L}) = \mathrm{Hom}_r(\mathcal{L})$. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are symmetric, then so is $\mathcal{L}_1 \wedge \mathcal{L}_2$. If $\mathcal{L}_1$ is symmetric, then so is $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ for each family $\mathcal{L}_2$.

From the definition, it is trivial that $\mathcal{L}_1 \wedge \mathcal{L}_2 = \mathcal{L}_2 \wedge \mathcal{L}_1$ and $(\mathcal{L}_1 \wedge \mathcal{L}_2) \wedge \mathcal{L}_3 = \mathcal{L}_1 \wedge (\mathcal{L}_2 \wedge \mathcal{L}_3)$. However, $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ need not equal $\mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_1)$. For example, if $\mathcal{L}_1 = \{\Sigma_1^2\}$ and $\mathcal{L}_2 = \{L_2\}$, where $L_2 = \{a^3/a \text{ in } \Sigma_1\}$, then $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) = \{\{(ab)^3/a, b \text{ in } \Sigma_1\}\} \neq \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_1) = \{\{a^3b^3/a, b \text{ in } \Sigma_1\}\}$. However, substitution does have the following associative properties:

<u>Proposition 1.1</u>. Let $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ be families of languages. Then

(a) $\mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3)) \subseteq \mathrm{Sub}(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_3)$.

(b) $\mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3)) = \mathrm{Sub}(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_3)$ if $\mathcal{L}_2$ is symmetric.

---

(3) A mapping h of Σ_1^* into Σ_2^* is called a <u>homomorphism</u> if $h(xy) = h(x)h(y)$ for all x and y in Σ_1^*.

(4) A homomorphism h on L is <u>restricted</u> on L if $h(w) = \epsilon$ for w in L implies $w = \epsilon$ and there is a positive integer q such that $h(w) \neq \epsilon$ for each subword w of length $\geq q$ of each word in L.

(5) The ordering, of course, is understood to be by family inclusion.

Proof. (a) Given L_3 in $\mathcal{L}_3$, let τ_2 be a substitution of L_3 by languages of $\mathcal{L}_2$[6] and τ_1 a substitution of $\tau_2(L_3)$ by languages of $\mathcal{L}_1$. Then $\tau_1(\tau_2(L_3)) = \tau'(L_3)$, where τ' is the substitution such that $\tau'(a) = \tau_1(\tau_2(a))$ for each a. Since $\tau_2(a)$ is in $\mathcal{L}_2$ and τ_1 is a substitution by languages of $\mathcal{L}_1$, $\tau_1(\tau_2(a))$ is in $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$. Therefore τ' is a substitution by languages of $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$, so that $\mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3)) \subseteq \mathrm{Sub}(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_3)$.

(b) Suppose $\mathcal{L}_2$ is symmetric. Let L_3 be in $\mathcal{L}_3$ and τ' a substitution of L_3 by languages of $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$. Then for each a in Σ_{L_3}, $\tau'(a) = \tau_{1,a}(L_{2,a})$, where $L_{2,a}$ is in $\mathcal{L}_2$ and $\tau_{1,a}$ is a substitution on $\Sigma_{L_{2,a}}$ by languages of $\mathcal{L}_1$. Since $\mathcal{L}_2$ is symmetric and Σ is infinite, we may assume that $\Sigma_{L_{2,a}} \cap \Sigma_{L_{2,a'}} = \emptyset$ for $a \neq a'$. Let $\Sigma_4 = \bigcup_a \Sigma_{L_{2,a}}$. Then there exists a substitution τ_1 of Σ_4^* by languages of $\mathcal{L}_1$ such that $\tau_1(b) = \tau_{1,a}(b)$ for each a in Σ_{L_3} and b in $\Sigma_{L_{2,a}}$. Hence $\tau'(L_3) = \tau_1(\tau_2(L_3))$, where τ_2 is the substitution of L_3 by languages of $\mathcal{L}_2$ defined by $\tau_2(a) = L_{2,a}$ for each a in Σ_{L_3}. Thus τ' is in $\mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3))$. Therefore $\mathrm{Sub}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_3] \subseteq \mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3)]$, whence equality by (a).

Remark. The hypothesis in (b) cannot always be removed and the result be true. For example, let $\Sigma = \{a_i/i\geq 1\}$, $\mathcal{L}_1 = \{\{a_i\}/i\geq 1\}$, $\mathcal{L}_2 = \{\{a_1\}\}$, and $\mathcal{L}_3 = \{\{a_i a_j\}/i, j\geq 1\}$. Then $\mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3) = \{\{a_1^2\}\}$, $\mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3)) = \{\{a_i^2\}/i\geq 1\}$, $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_1$, and $\mathrm{Sub}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_3] = \mathcal{L}_3$. Clearly $\mathrm{Sub}(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_3)$ is not contained in $\mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{L}_3))$.

Recently, families of languages with six additional properties have been introduced [3] because of their intimate connection with families of languages

[6] A substitution τ is a substitution by languages of $\mathcal{L}_2$ if $\tau(a)$ is in $\mathcal{L}_2$ for each a.

of interest in computer science. These families, called "abstract families of languages" are currently under extensive investigation and the present paper may be viewed as an addition to their literature.

Definition. As abstract family of languages (abbreviated AFL) is a family of languages closed under union, concatenation, + [7], ϵ-free homomorphism [8], inverse homomorphism [9], and intersection with regular sets. [10]

If h is a homomorphism from Σ_1^* into Σ_2^* then h^{-1}, the inverse homomorphism, is the mapping from $2^{\Sigma_2^*}$ into $2^{\Sigma_1^*}$ defined by $h^{-1}(A) = \{w/h(w) \text{ in } A\}$ for each $A \subseteq \Sigma_2^*$.

For our purposes it is convenient to consider a reformulation of the closure properties of an AFL, expressed as follows:

Notation. Let $\mathcal{R}$ be the family of regular sets (over Σ) and $\mathcal{R}_0$ the family of ϵ-free regular sets.

Proposition 1.2. A family $\mathcal{L}$ of languages is an AFL if and only if (1) $\mathcal{R}_0 \subseteq \mathcal{L}$, (2) $\mathrm{Sub}(\mathcal{R}_0, \mathcal{L}) \subseteq \mathcal{L}$, (3) $\mathrm{Sub}(\mathcal{L}, \mathcal{R}_0) \subseteq \mathcal{L}$, (4) $\mathcal{L} \wedge \mathcal{R} \subseteq \mathcal{L}$, and (5) $\mathrm{Hom}_r(\mathcal{L}) \subseteq \mathcal{L}$.

Proof. It is known [3] that each AFL satisfies all five of the conditions. Thus consider the converse. Assume $\mathcal{L}$ satisfies (1) - (5). Thus $\mathcal{L}$ has all the closure properties of an AFL except possibly for closure under inverse homomorphism. Suppose $\mathcal{L}$ contains a language containing ϵ. It follows from (4) that $\mathcal{L}$ contains $\{\epsilon\}$ and, from (1) and the fact that $\mathcal{L}$ is closed under union, that $\mathcal{L}$ contains $\mathcal{R}$.

---

(7) For each set $A \subseteq \Sigma^*$, $A^+ = \bigcup_{i \geq 1} A^i$.

(8) A homomorphism h is called ϵ-free if $h(w) = \epsilon$ implies $w = \epsilon$.

(9) If h is a homomorphism from Σ_1^* into Σ_2^*, then h^{-1}, the inverse homomorphism, is the mapping from $2^{\Sigma_2^*}$ into $2^{\Sigma_1^*}$ defined by $h^{-1}(A) = \{w/h(w) \text{ in } A\}$ for each $A \subseteq \Sigma_2^*$.

(10) The family of regular sets is the smallest family of languages containing all the finite languages and closed under union, concatenation, and *, where $A^* = A^+ \cup \{\epsilon\}$ for each $A \subseteq \Sigma^*$.

It then follows from Theorem 4 of [7] that $\mathcal{L}$ is closed under inverse homomorphism. Suppose $\mathcal{L}$ is ϵ-free. By an argument similar to the one given in the proof of Theorem 4 of [7], it follows that $\mathcal{L}$ is closed under inverse homomorphism. In either case, therefore, $\mathcal{L}$ is closed under inverse homomorphism and so is an AFL.

<u>Remark</u>. In the presence of conditions (2) - (5), condition (1) is equivalent to the condition that $\mathcal{L}$ is nontrivial. That is, the only family $\mathcal{L}$ satisfying (2) - (5) but not (1) is the family $\mathcal{L} = \{\{\epsilon\}, \emptyset\}$.

<u>Notation</u>. Given a family $\mathcal{L}$ let AFL($\mathcal{L}$) be the smallest AFL containing $\mathcal{L}$.

It is known [3] that AFL($\mathcal{L}$) exists for each family $\mathcal{L}$.

<u>Corollary</u>. Given a family $\mathcal{L}$, let $\mathcal{L}^{(o)} = \mathcal{L} \cup R_o$ and for each $n \geq 0$ let $\mathcal{L}^{(n+1)} =$ $\mathrm{Sub}(R_o, \mathcal{L}^{(n)})$ if $n \equiv 0 \pmod 4$, $\mathcal{L}^{(n+1)} = \mathrm{Sub}(\mathcal{L}^{(n)}, R_o)$ if $n \equiv 1 \pmod 4$, $\mathcal{L}^{(n+1)} =$ $\mathcal{L}^{(n)} \wedge R$ if $n \equiv 2 \pmod 4$, and $\mathcal{L}^{(n+1)} = \mathrm{Hom}_r(\mathcal{L}^{(n)})$ if $n \equiv 3 \pmod 4$. Then $\mathcal{L} \subseteq \mathcal{L}^{(o)} \subseteq \ldots \subseteq \mathcal{L}^{(n)} \subseteq \ldots$ and AFL $(\mathcal{L}) = \bigcup_{i \geq o} \mathcal{L}^{(i)}$.

<u>Proof</u>. Let $\mathcal{L}'$ be any AFL containing $\mathcal{L}$. Then $\mathcal{L}^{(o)} \subseteq \mathcal{L}'$ and a simple induction on n shows that $\mathcal{L}^{(n)} \subseteq \mathcal{L}'$ for all $n \geq 0$. Therefore $\bigcup_{n \geq o} \mathcal{L}^{(n)} \subseteq \mathcal{L}'$. To complete the proof it suffices to verify that $\bigcup_{n \geq o} \mathcal{L}^{(n)}$ is an AFL.

Clearly $\bigcup_{n \geq o} \mathcal{L}^{(n)}$ satisfies condition (1) of Proposition 1.2.

Since

$$\mathrm{Sub}(R_o, \bigcup_{n \geq o} \mathcal{L}^{(n)}) = \mathrm{Sub}(R_o, \bigcup_{n \equiv o} \mathcal{L}^{(n)}) = \bigcup_{n \equiv o} \mathcal{L}^{(n+1)} = \bigcup_{n \geq o} \mathcal{L}^{(n)},$$

$\bigcup_n \mathcal{L}^{(n)}$ satisfies condition (2). A similar calculation shows that it satisfies conditions (4) and (5). Finally, observe that each substitution τ by languages of $\bigcup_{n \geq o} \mathcal{L}^{(n)}$ involves only finitely many languages. Therefore there is some m such that τ is a substitution by languages of $\mathcal{L}^{(m)}$. Hence

$$\mathrm{Sub}\Big(\bigcup_{n\geq 0} \mathcal{L}^{(n)}, \mathcal{R}_o\Big) \subseteq \bigcup_{n\geq 0} \mathrm{Sub}(\mathcal{L}^{(n)}, \mathcal{R}_o) = \bigcup_{n\equiv 1} \mathrm{Sub}(\mathcal{L}^{(n)}, \mathcal{R}_o)$$

$$= \bigcup_{n\equiv 1} \mathcal{L}^{(n+1)} = \bigcup_{n\geq 0} \mathcal{L}^{(n)},$$

whence condition (3). Therefore $\bigcup_{n\geq 0} \mathcal{L}^{(n)}$ is an AFL.

A particularly important type of AFL is one which is closed under arbitrary homomorphism. That is,

Definition. An AFL $\mathcal{L}$ is said to be full [3] if it is closed under arbitrary homomorphism.

The following result provides a useful reformulation of full AFL.

Proposition 1.3. A family $\mathcal{L}$ of languages is a full AFL if and only if (1) $\mathcal{R}_o \subseteq \mathcal{L}$, (2) $\mathrm{Sub}(\mathcal{R}, \mathcal{L}) \subseteq \mathcal{L}$, (3) $\mathrm{Sub}(\mathcal{L}, \mathcal{R}_o) \subseteq \mathcal{L}$, and (4) $\mathcal{L} \wedge \mathcal{R} \subseteq \mathcal{L}$.

Proof. Suppose $\mathcal{L}$ is a full AFL. By Proposition 1.2, $\mathcal{L}$ satisfies (1), (3), and (4). By Theorem 2.4 of [3], $\mathcal{L}$ satisfies (2).

Now assume $\mathcal{L}$ satisfies the above conditions. Then $\mathcal{L}$ satisfies condition (1) - (4) of Proposition 1.2. Since each homomorphism is a substitution by languages of $\mathcal{R}$, (2) above implies $\mathcal{L}$ is closed under homomorphism. Thus $\mathcal{L}$ satisfies (1) - (5) of Proposition 1.2 and so is an AFL. Being closed under arbitrary homomorphism, $\mathcal{L}$ is a full AFL.

Notation. For each AFL $\mathcal{L}$, let $AFL_f(\mathcal{L})$ be the smallest full AFL containing $\mathcal{L}$.

Corollary. Given a family $\mathcal{L}$, let $\mathcal{L}^{(o)} = \mathcal{L} \cup \mathcal{R}_o$ and for each $n \geq o$, let $\mathcal{L}^{(n+1)} = \mathrm{Sub}(\mathcal{R}, \mathcal{L}^{(n)})$ if $n \equiv o \pmod 3$, $\mathcal{L}^{(n+1)} = \mathrm{Sub}(\mathcal{L}^{(n)}, \mathcal{R}_o)$ if $n \equiv 1 \pmod 3$, and $\mathcal{L}^{(n+1)} = \mathcal{L}^{(n)} \wedge \mathcal{R}$ if $n \equiv 2 \pmod 3$. Then $\mathcal{L} \subseteq \mathcal{L}^{(o)} \subseteq \ldots \subseteq \mathcal{L}^{(n)} \subseteq \ldots$ and $AFL_f(\mathcal{L}) = \bigcup_{n\geq o} \mathcal{L}^{(n)}$.

The proof is analogous to that of the corollary to Proposition 1.2 and is omitted.

Section 2. Substitution of AFL

In this section we show that the substitution of an AFL into an AFL is an AFL. To do this we first prove two technical lemmas. The first asserts a kind of distributivity of regular set intersection with respect to substitution and the second a kind of distributivity of Hom_r with respect to substitution. (These lemmas are also used in section 4.)

Notation. Let $\mathcal{F}_o$ be the family of ϵ-free finite sets.

Note that $\mathcal{F}_o$ is symmetric and $\mathrm{Sub}(\mathcal{F}_o, \mathcal{F}_o) = \mathcal{F}_o$.

Lemma 2.1. For all families of languages $\mathcal{L}_1$ and $\mathcal{L}_2$,

$$\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \wedge \mathcal{R} \subseteq \mathrm{Sub}(\mathcal{L}_1 \wedge \mathcal{R}, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2) \wedge \mathcal{R}).$$

Proof. Let $\mathcal{L}_2' = \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2)$. Then $\mathcal{L}_2 \subseteq \mathcal{L}_2'$ and $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \wedge \mathcal{R} \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2') \wedge \mathcal{R}$. By (b) of Proposition 1.1, $\mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2') = \mathrm{Sub}(\mathcal{F}_o, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2)) = \mathrm{Sub}(\mathrm{Sub}(\mathcal{F}_o, \mathcal{F}_o), \mathcal{L}_2) = \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2)$. Thus

$$\mathrm{Sub}(\mathcal{L}_1 \wedge \mathcal{R}, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2') \wedge \mathcal{R}) = \mathrm{Sub}(\mathcal{L}_1 \wedge \mathcal{R}, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2) \wedge \mathcal{R}).$$

Hence it suffices to show the lemma for $\mathcal{L}_2$ replaced by $\mathcal{L}_2'$. Thus, without loss of generality, we may assume that $\mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2) = \mathcal{L}_2$. Using this assumption it suffices to show

$$\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \wedge \mathcal{R} \subseteq \mathrm{Sub}(\mathcal{L} \wedge \mathcal{R}, \mathcal{L}_2 \wedge \mathcal{R}.)$$

Since $\mathcal{L}_2$ is closed under substitution by $\mathcal{F}_o$, it follows that $\mathcal{L}_2$ is symmetric. Let L_2 be in $\mathcal{L}_2$ and let τ_1 be a substitution of L_2 by languages in $\mathcal{L}_1$. Let R be a regular set, with $R \subseteq \Sigma_1^*$. By extending Σ_1 if necessary, we may assume that $\tau_1(L_2) \subseteq \Sigma_1^*$. Since $\mathcal{L}_2$ is symmetric, we may assume that $L_2 \subseteq \Sigma_2^*$, with $\Sigma_1 \cap \Sigma_2 = \emptyset$. Let τ_2 be the substitution on Σ_2^* defined by $\tau_3(a) = a\,\tau_1(a)$ for each a in Σ_2. Let h be the homomorphism on $(\Sigma_1 \cup \Sigma_2)^*$ defined by $h(a) = \epsilon$ for each a

in Σ_2 and $h(b) = b$ for each b in Σ_1. Then $\tau_1 = h\tau_2$ and

$$\tau_1(L_2)\cap R = h\tau_2(L_2)\cap R = h[\tau_2(L_2)\cap h^{-1}(R)].$$

Since R is regular, $h^{-1}(R)$ is regular [5]. Thus there exists an fsa(11) $A = (K, \Sigma_1 \cup \Sigma_2, \delta, p_o, F)$ such that (12) $T(A) = h^{-1}(R)$. Let

$$R' = (T(A)\cap\{\epsilon\}) \cup \{(a_1,p_o,p_1)\ \dots\ (a_m,p_{m-1},p_m)/m\geq 1, \text{ each } a_i \text{ in } \Sigma_2, \text{ each } p_i \text{ in } K, \text{ and } p_m \text{ in } F\}.$$

As is well known [2], R' is regular.

For each (a,p,q) in $\Sigma_2 \times K \times K$, let

$$R(a,p,q) = \{w \text{ in } \Sigma_1^*/\delta(p,aw) = q\}.$$

Then $R(a,p,q)$ is regular (since $R(a,p,q) = T(B)$, where B is the fsa $(K, \Sigma_1, \delta, \delta(p,a), \{q\})$). Let τ_3 be the substitution defined on $(\Sigma_2 \times K \times K)^*$ by $\tau_3((a,p,q)) = a\,R(a,p,q)$ for each (a,p,q) in $\Sigma_2 \times K \times K$. Then $\tau_3(R')$ is the subset of $h^{-1}(R)$ of all words that do not begin with a symbol of Σ_1. Since $\tau_2(L_2)$ contains no word beginning with a symbol of Σ_1,

$$\tau_2(L_2) \cap h^{-1}(R) = \tau_2(L_2) \cap \tau_3(R').$$

Let τ' be the substitution on Σ_2^* defined by $\tau'(a) = \{a\}\times K \times K$ for each a in Σ_2 and τ'' the substitution on $(\Sigma_2 \times K \times K)^*$ by $\tau''((a,p,q)) = \tau_2(a)$ for each (a,p,q). Then $\tau_2 = \tau''\tau'$, so that

$$\tau_2(L_2) \cap h^{-1}(R) = \tau''\tau'(L_2) \cap \tau_3(R').$$

---

(11) As fsa (<u>finite state acceptor</u>) is a 5-tuple $A = (K, \Sigma_1, \delta, p_o, F)$, where (i) K and Σ_1 are finite sets (of <u>states</u> and <u>inputs</u>, resp.), (ii) δ is a function from $K\times\Sigma_1$ to K (the <u>next state</u> function), (iii) s_o is an element of K (the <u>start</u> state), and (v) $F \subseteq K$ (the set of <u>accepting</u> states). The function δ is extended inductively to $K \times \Sigma_1^*$ by letting $\delta(p,\epsilon) = p$ and $\delta(p,a_1 \dots a_n) = \delta[\delta(p,a_1 \dots a_{n-1}),a_n]$ for each p in K, each $n\geq 1$, and each $a_1, \dots, a_n$ in Σ_1.

(12) For each fsa A, $T(A) = \{w \text{ in } \Sigma_1^*/\delta(p_o,w) \text{ in } F\}$. It is known [8] that a set $R \subseteq \Sigma_1^*$ is regular if and only if $T(A) = R$ for some fsa A.

Let τ_3'' be the substitution on $(\Sigma_2 \times K \times K)^*$ defined by

$$\tau_3''((a,p,q)) = \tau''((a,p,q)) \cap \tau_3((a,p,q))$$
$$= a\tau_1(a) \cap aR(a,p,q)$$

for each (a,p,q).

We now show that

$$(*)\ \tau''\tau'(L_2) \cap \tau_3(R') = \tau_3''[\tau'(L_2) \cap R'].$$

Since $\tau_3''((a,p,q)) = \tau''((a,p,q)) \cap \tau_3((a,p,q))$, we have $\tau_3''(R') \subseteq \tau_3(R')$ and $\tau_3''[\tau'(L_2)] \subseteq \tau''[\tau'(L_2)]$. Thus

$$\tau_3''[\tau'(L_2) \cap R'] \subseteq \tau_3''\ \tau'(L_2) \cap \tau_3''(R')$$
$$\subseteq \tau''\ \tau'(L_2) \cap \tau_3(R').$$

To see the reverse containment, assume w is in $\tau''\tau'(L_2) \cap \tau_3(R')$. Suppose $w = \epsilon$. Then ϵ is in L_2 and in R', since $\tau''\tau'$ and τ_3 are ϵ-free substitutions.(13) Therefore ϵ is in $\tau_3''[\tau'(L_2) \cap R']$. Suppose $w \neq \epsilon$. Then $w = a_1x_1 \ldots a_qx_q$ for some $q \geq 1$, some $a_1, \ldots, a_q$ in Σ_2, and some $x_1, \ldots, x_q$ in Σ_1^*. Since w is in $\tau''\tau'(L_2) = \tau_2(L_2)$, it follows that $a_1 \ldots a_q$ is in L_2 and x_i is in $\tau_1(a_i)$ for each i. Define p_i in K by induction on i so that $p_{j+1} = \delta(p_j, a_{j+1}x_{j+1})$ for each j, $0 \leq j < q$. Let $w_2 = (a_1,p_0,p_1) \ldots (a_q,p_{q-1},p_q)$. Since w is in $\tau_3(R')$, w_2 is in R'. Since $\tau'(a) = \{a\} \times K \times K$ for each a, w_2 is in $\tau'(a_1 \ldots a_q)$. Therefore w_2 is in $\tau'(L_2) \cap R'$. By definition of τ_3'', a_ix_i is in $\tau_3''(a_i,p_{i-1},p_i)$ for $1 \leq i \leq q$. Therefore, w is in $\tau_3''(w_2)$ so that w is in $\tau_3''[\tau'(L_2) \cap R']$. Hence $\tau''\tau'(L_2) \cap \tau_3(R') \subseteq \tau''[\tau'(L_2) \cap R']$, implying $(*)$.

---

(13) A substitution τ is <u>ϵ-free</u> if ϵ in $\tau(a)$ implies $a = \epsilon$.

From (*), it follows that

$$\begin{aligned}\tau_1(L_2) \cap R &= h[\tau_2(L_2) \cap h^{-1}(R)] \\ &= h[\tau''\tau'(L_2) \cap \tau_3(R')] \\ &= h\tau_3''[\tau'(L_2) \cap R'].\end{aligned}$$

Now $\tau'(L_2)$ is in $\mathrm{Sub}(\mathcal{F}_0, \mathcal{L}_2) = \mathcal{L}_2$ and $\tau'(L_2) \cap R'$ is in $\mathcal{L}_2 \wedge \mathcal{R}$. The composite $h\tau_3''$ is a substitution such that $h\tau_3''((a,p,q)) = h[a\tau_1(a) \cap aR(a,p,q)] = \tau_1(a) \cap R(a,p,q)$ for each (a,p,q). Since $\tau_1(a)$ is in $\mathcal{L}_1$ and $R(a,p,q)$ in $\mathcal{R}$, $h\tau_3''$ is a substitution by sets of $\mathcal{L}_1 \wedge \mathcal{R}$. Therefore $\tau_1(L_2) \cap R$ is in $\mathrm{Sub}(\mathcal{L}_1 \wedge \mathcal{R}, \mathcal{L}_2 \wedge \mathcal{R})$, and the lemma is proved.

The result pertaining to the distributivity of Hom_r is

Lemma 2.2. For all families of languages $\mathcal{L}_1$ and $\mathcal{L}_2$,

$$\mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] \subseteq \mathrm{Sub}[\mathrm{Hom}_r(\mathcal{L}_1 \wedge \mathcal{R}), \mathrm{Hom}_r(\mathrm{Sub}(\mathcal{F}_0, \mathcal{L}_2))].$$

Proof. Let L_2 be a language in $\mathcal{L}_2$, τ a substitution of L_2 by languages of $\mathcal{L}_1$, $\Sigma_1 = \Sigma_{\tau(L_2)}$, $\Sigma_2 = \Sigma_{L_2}$, and h a homomorphism from Σ_1^* to Σ_3^* which is restricted on $\tau(L_2)$. By definition of Σ_2, for each a in Σ_2, there is a word w in L_2 containing an occurrence of a. Let R be the set containing ϵ and all words w in Σ_1^* such that $h(w) \neq \epsilon$. Then $R = \{\epsilon\} \cup h^{-1}(\Sigma_3^+)$ is a regular set. Since, for each a in Σ_2, each subword of a word of $\tau(a)$ is also a subword of some word of $\tau(L_2)$, h is restricted on $\tau(a) \cap R$. Let τ'' be the substitution by languages of $\mathrm{Hom}_r(\mathcal{L}_1 \wedge \mathcal{R})$ defined by $\tau''(a) = h[\tau(a) \cap R]$ for each a in Σ_2.

For each a in Σ_2 let a' be a new symbol and let $\Sigma_2' = \{a'/a \text{ in } \Sigma_2\}$. Let τ' be the substitution on Σ_2^* defined by $\tau'(a) = \{a\}$ if $\tau(a)$ contains ϵ or $h\tau(a)$ is ϵ-free, and $\tau'(a) = \{a, a'\}$ if $\tau(a)$ is ϵ-free and $h\tau(a)$ contains ϵ. Then τ' is a substitution by languages of $\mathcal{F}_0$. Let h' be the homomorphism on $(\Sigma_2 \cup \Sigma_2')^*$ defined

by $h'(a) = a$ and $h'(a') = \epsilon$ for each a in Σ_2. Then h' is restricted on $\tau'(L_2)$. [For let $\Sigma_4 = \{a/\tau'(a) = a\}$ and $\Sigma_5 = \{a/\tau'(a) = \{a, a'\}\}$. By definition of τ', for each a in Σ_5 there is a non-ϵ word w_a in $\tau(a)$ such that $h(w_a) = \epsilon$. Since h is restricted on $\tau(L_2)$ it follows that (α) $L_2 \cap \Sigma_5^* = L_2 \cap \{\epsilon\}$ and (β) there exists $q \geq 0$ such that any subword of length $> q$ of a word of L_2 contains an occurrence of an element of Σ_4. Now (α) implies that if $h'(w') = \epsilon$ for w' in $\tau'(L_2)$ then $w' = \epsilon$, and (β) implies that if $h'(u') = \epsilon$ for a subword u' of some word of $\tau'(L_2)$ then $|u'| \leq q$. Therefore h' is restricted on $\tau'(L_2)$.] Therefore $h'\tau'(L_2)$ is a language in $\mathrm{Hom}_r(\mathrm{Sub}(\mathcal{F}_0, \mathcal{L}_2))$, so that $\tau''h'\tau'(L_2)$ is in $\mathrm{Sub}[\mathrm{Hom}_r(\mathcal{L}_1 \wedge R), \mathrm{Hom}_r(\mathrm{Sub}(\mathcal{F}_0, \mathcal{L}_2))]$.

To complete the proof it suffices to show that $h\tau(L_2) = \tau''h'\tau'(L_2)$. Therefore let w be a word in L_2. We shall show that $h\tau(w) = \tau''h'\tau'(w)$. If $w = \epsilon$, then $h\tau(w) = \epsilon = \tau''h'\tau'(w)$. Assume $w = a_1 \dots a_n$, $n \geq 1$, each a_i in Σ_2. By definition,

$$h\tau(a_1 \dots a_n) = \{h(w_1) \dots h(w_n)/w_i \text{ in } \tau(a_i)\}.$$

For each i, let w_i be a word of $\tau(a_i)$. Let J be the set of all j such that $h(w_j) = \epsilon$ and $\tau(a_j)$ is ϵ-free, and let $J' = \{1, \dots, n\} - J$. Then $h(w_{j'})$ is in $\tau''(a_{j'})$ for each j' in J'. For each i, $1 \leq i \leq n$, let $b_i = a_i'$ if i is in J and $b_i = a_i$ if i is in J'. Since $h(w_j) = \epsilon$ and $\tau(a_j)$ is ϵ-free for each j in J, $b_1 \dots b_n$ is in $\tau'(a_1 \dots a_n)$. If i is in J then $h'(b_i) = \epsilon = h(w_i)$ and $h(w_i)$ is in $\tau''h'(b_i)$. If i is in J', then $h'(b_i) = a_i$ and $h(w_i)$ is in $\tau''(a_i)$. Thus $h(w_1) \dots h(w_n)$ is in $\tau''(h'(b_1 \dots b_n))$, so that $h\tau(w) \subseteq \tau''h'\tau'(w)$.

To see the reverse containment, note that

$$\tau''h'\tau'(a_1 \dots a_n) = \{u_1 \dots u_n/u_i \text{ in } \tau''h'(b_i),\ b_i \text{ in } \tau'(a_i)\}.$$

For each i, if $b_i = a_i'$ then $h'(b_i) = \epsilon$, $u_i = \epsilon$, $\tau(a_i)$ is ϵ-free, and $h\tau(a_i)$ contains $\epsilon = u_i$; and if $b_i = a_i$, then u_i is in $\tau''(a_i) = h[\tau(a_i) \cap R] \subseteq h\tau(a_i)$. In either case, u_i is in $h\tau(a_i)$. Thus $u_1 \ldots u_n$ is in $h\tau(a_1 \ldots a_n)$ and $\tau''h'\tau'(w) \subseteq h\tau(w)$.

We are now ready for the main result of the section.

<u>Theorem 2.1</u>. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are AFL, then so is $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$.

<u>Proof</u>. Since $\mathcal{R}_o \subseteq \mathcal{L}_1$ and $\mathcal{R}_o \subseteq \mathcal{L}_2$, it follows that $\mathcal{R}_o = \mathrm{Sub}(\mathcal{R}_o, \mathcal{R}_o) \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$. Thus $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ satisfies (1) of Proposition 1.2. By Propositions 1.1 and 1.2,

$$\mathrm{Sub}(\mathcal{R}_o, \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)) = \mathrm{Sub}(\mathrm{Sub}(\mathcal{R}_o, \mathcal{L}_1), \mathcal{L}_2) \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$$

and

$$\mathrm{Sub}(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{R}_o) = \mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}_2, \mathcal{R}_o)) \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2).$$

Thus $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ satisfies conditions (2) and (3) of Proposition 1.2. By Lemma 2.1,

$$\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \wedge \mathcal{R} \subseteq \mathrm{Sub}(\mathcal{L}_1 \wedge \mathcal{R}, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2) \wedge \mathcal{R}) \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2).$$

By Lemma 2.2,

$$\mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] \subseteq \mathrm{Sub}[\mathrm{Hom}_r(\mathcal{L}_1 \wedge \mathcal{R}), \mathrm{Hom}_r(\mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2))] \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2).$$

Thus $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ satisfies conditions (4) and (5) of Proposition 1.2. Hence $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ is an AFL.

<u>Corollary 1</u>. If $\mathcal{L}_1$ is a full AFL and $\mathcal{L}_2$ is an AFL, then $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ is a full AFL.

<u>Proof</u>. By Theorem 2.1, $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ is an AFL. By Propositions 1.1 and 1.3, we have

$$\mathrm{Sub}(\mathcal{R}, \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)) = \mathrm{Sub}(\mathrm{Sub}(\mathcal{R}, \mathcal{L}_1), \mathcal{L}_2) \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2).$$

Thus $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ satisfies (1) - (4) of Proposition 1.3 and so is a full AFL.

Corollary 2. If $\mathcal{L}$ is an AFL, then $AFL_f(\mathcal{L}) = Sub(\mathcal{R}, \mathcal{L})$.

Proof. By Corollary 1, $Sub(\mathcal{R}, \mathcal{L})$ is a full AFL. If $\mathcal{L}'$ is any full AFL containing $\mathcal{L}$, then $Sub(\mathcal{R}, \mathcal{L}) \subseteq Sub(\mathcal{R}, \mathcal{L}') \subseteq \mathcal{L}'$ by (2) of Proposition 1.3. Thus $Sub(\mathcal{R}, \mathcal{L})$ is the smallest full AFL containing $\mathcal{L}$, i.e., $Sub(\mathcal{R}, \mathcal{L}) = AFL_f(\mathcal{L})$.

Remarks (1) Theorem 2.1 and Corollary 1 both hold if $\mathcal{L}_2$ is a family of languages such that $Sub(\mathcal{R}_o, \mathcal{L}_2)$ is an AFL. For in this case

$$Sub(\mathcal{L}_1, \mathcal{L}_2) \subseteq Sub(\mathcal{L}_1, Sub(\mathcal{R}_o, \mathcal{L}_2)) = Sub(Sub(\mathcal{L}_1, \mathcal{R}_o), \mathcal{L}_2) \subseteq Sub(\mathcal{L}_1, \mathcal{L}_2).$$

Thus $Sub(\mathcal{L}_1, \mathcal{L}_2) = Sub(\mathcal{L}_1, Sub(\mathcal{R}_o, \mathcal{L}_2))$ is a (full) AFL if $\mathcal{L}_1$ is a (full) AFL. In particular, the results are valid if $\mathcal{L}_2$ is a pre-AFL or ϵ-free pre-AFL [4]. Similarly, Corollary 2 is valid if $\mathcal{L}$ is any family of languages such that $Sub(\mathcal{R}_o, \mathcal{L})$ is an AFL, hence if $\mathcal{L}$ is a pre-AFL or ϵ-free pre-AFL.

(2) Theorem 2.1 suggests the following general problem (which is not studied here): "Identify" $Sub(\mathcal{L}_1, \mathcal{L}_2)$ for well-known AFL $\mathcal{L}_1$ and $\mathcal{L}_2$.

Section 3. AFL of Substitutions

This section is concerned with relations between substitution and the AFL generated by a family. The main result asserts that, with suitable hypotheses,

$$AFL[Sub(\mathcal{L}_1, \mathcal{L}_2)] = Sub[AFL(\mathcal{L}_1), AFL(\mathcal{L}_2)].$$

In order to prove the main result, a sequence of lemmas is needed.

Lemma 3.1. Let $\mathcal{L}_1$ be an ϵ-free (or arbitrary) family of languages. Then for every family of languages $\mathcal{L}_2$,

$$Sub(\mathcal{L}_1 \wedge \mathcal{R}, \mathcal{L}_2 \wedge \mathcal{R}) \subseteq Hom_r[Sub(\mathcal{L}_1 \cup \mathcal{F}_o, Sub(\mathcal{F}_o, \mathcal{L}_2)) \wedge \mathcal{R}]$$

(or $Sub(\mathcal{L}_1 \wedge \mathcal{R}, \mathcal{L}_2 \wedge \mathcal{R}) \subseteq Hom[Sub(\mathcal{L}_1 \cup \mathcal{F}_o, Sub(\mathcal{F}_o, \mathcal{L}_2)) \wedge \mathcal{R}]$).

Proof. We prove the lemma for the ϵ-free case, the argument for arbitrary $\mathcal{L}_1$ differing trivially.

Let L_2 be in $\mathcal{L}_2$, R_2 in $\mathcal{R}$, and consider $L_2 \cap R_2$, with $L_2 \cup R_2 \subseteq \Sigma_2^*$. Let τ be a substitution on Σ_2^* by languages of $\mathcal{L}_1 \wedge \mathcal{R}$. Then for each a in Σ_2, $\tau(a) = L_a \cap R_a$, where L_a is in $\mathcal{L}_1$ and R_a is in $\mathcal{R}$. For each a in Σ_2 let a' be a new symbol and let $\Sigma_2' = \{a'/a \text{ in } \Sigma_2\}$. Let τ' be the substitution on Σ_2^* defined by $\tau'(a) = \{a'a\}$ for each a. Then τ' is a substitution by languages of $\mathcal{F}_0$, so that $\tau'(L_2)$ is in Sub$(\mathcal{F}_0, \mathcal{L}_2)$. Let τ'' be the substitution on $(\Sigma_2 \cup \Sigma_2')^*$ defined by $\tau''(a') = \{a'\}$ and $\tau''(a) = L_a$ for each a in Σ_2. Then τ'' is a substitution by languages of $\mathcal{L}_1 \cup \mathcal{F}_0$, so that $\tau''\tau'(L_2)$ is in Sub$[\mathcal{L}_1 \cup \mathcal{F}_0$, Sub$(\mathcal{F}_0, \mathcal{L}_2)]$. Let $R' = (\bigcup_{a \text{ in } \Sigma_2} a'R_a)^*$ and let $R'' = \tau'''(R_2)$, where τ''' is the substitution on Σ_2^* defined by $\tau'''(a) = a' \Sigma_{L_a}^*$ for each a. Then R' and R'' are regular sets,(14) and $\tau''\tau'(L_2) \cap R' \cap R''$ is in Sub$[\mathcal{L}_1 \cup \mathcal{F}_0$, Sub$(\mathcal{F}_0, \mathcal{L}_2)] \wedge \mathcal{R}$. Let h be the homomorphism such that $h(a') = \epsilon$ for each a' and $h(b) = b$ if b is a symbol not in Σ_2'. Since $\mathcal{L}_1$ is ϵ-free, h cannot erase two consecutive symbols of a word in $\tau''\tau'(L_2)$. Furthermore, if w is in $\tau''\tau'(L_2)$ and $h(w) = \epsilon$, then $w = \epsilon$. Therefore h is restricted on $\tau''\tau'(L_2)$ and thus restricted on $\tau''\tau'(L_2) \cap R' \cap R''$.

To complete the proof it suffices to show that $\tau(L_2 \cap R_2) = h[\tau''\tau'(L_2) \cap R' \cap R'']$. Let w be any word in Σ_2^*. If $w = \epsilon$, then $\tau(\epsilon) = \{\epsilon\} = h[\tau''\tau'(\epsilon) \cap R']$. Suppose $w = a_1 \ldots a_n$, $n \geq 1$, each a_i in Σ_2. Then

$$\tau(w) = \{w_1 \ldots w_n / w_i \text{ in } L_{a_i} \cap R_{a_i}\}$$

$$= h\{a_1'w_1 \ldots a_n'w_n / w_i \text{ in } L_{a_i} \cap R_{a_i}\}$$

(14) Regular sets are closed under intersection [8] and under substitution by regular sets [1].

$$= h[\{a_1'u_1 \ldots a_n'u_n/u_i \text{ in } L_{a_i}\} \cap \{a_1'v_1 \ldots a_n'v_n/v_i \text{ in } R_{a_i}\}]$$
$$= h[\tau''\tau'(w) \cap R'].$$

Therefore $\tau(w) = h[\tau''\tau'(w) \cap R']$ for every word w in Σ_2^*. Hence $\tau(L_2 \cap R_2) = h[\tau''\tau'(L_2 \cap R_2) \cap R']$.

Clearly $\tau''\tau'(R_2) \subseteq R''$. Hence $\tau''\tau'(L_2 \cap R_2) \subseteq \tau''\tau'(L_2) \cap R''$. We prove the reverse inclusion, thereby obtaining equality.

Suppose w is in $\tau''\tau'(L_2) \cap R''$. Assume $w = \epsilon$. Then ϵ is in $\tau''\tau'(L_2)$ and in R'', thus in L_2 and in R_2. Hence ϵ is in $L_2 \cap R_2$ and therefore in $\tau''\tau'(L_2 \cap R_2)$. Assume $w = a_1'w_1 \ldots a_n'w_n$, $n \geq 1$, is in $\tau''\tau'(L_2) \cap R''$. Then each w_i is in L_{a_i}, and $a_1 \ldots a_n$ is in L_2 and in R_2. Therefore w is in $\tau''\tau'(L_2 \cap R_2)$. Thus $\tau''\tau'(L_2) \cap R'' \subseteq \tau''\tau'(L_2 \cap R_2)$.

Since $\tau''\tau'(L_2 \cap R_2) = \tau''\tau'(L_2) \cap R''$, we have

$$\tau(L_2 \cap R_2) = h[\tau''\tau'(L_2 \cap R_2) \cap R']$$
$$= h[\tau''\tau'(L_2) \cap R'' \cap R'].$$

<u>Lemma 3.2</u>. Let $\mathcal{L}_2$ be a family of languages. Then

(a) for each ϵ-free symmetric family $\mathcal{L}_1$,

$$\mathrm{Sub}[\mathrm{Hom}_r(\mathcal{L}_1), \mathcal{L}_2] \subseteq \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)].$$

(b) for each ϵ-free family $\mathcal{L}_1$,

$$\mathrm{Sub}[\mathcal{L}_1, \mathrm{Hom}_r(\mathcal{L}_2)] \subseteq \mathrm{Hom}_r(\mathrm{Sub}[\mathcal{L}_1 \cup \mathcal{F}_o, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2)]).$$

<u>Proof</u>.(a) Let L_2 be in $\mathcal{L}_2$ and τ a substitution of $\Sigma_{L_2}^*$ by languages of $\mathrm{Hom}_r(\mathcal{L}_1)$. Then for each a in Σ_{L_2}, $\tau(a) = h_a(L_a)$, where L_a is in $\mathcal{L}_1$ and h_a is restricted on L_a. Since $\mathcal{L}_1$ is symmetric, we may assume that $\Sigma_{L_a} \cap \Sigma_{L_b} = \emptyset$ for $a \neq b$. Then there is a homomorphism h on $(\cup_a \Sigma_{L_a})^*$ such that $h(x) = h_a(x)$, where x is in Σ_{L_a}. Hence τ is the composite $h\tau'$, where τ' is the substitution on $\Sigma_{L_2}^*$ defined by $\tau'(a) = L_a$ for each a. Since τ' is a substitution by languages of $\mathcal{L}_1$, it

suffices to show that h is restricted on $\tau'(L_2)$.

Since h_a is restricted on L_a for each a there exists $q_a>0$ such that $|w'|\le q_a$ whenever w' is a subword of a word of L_a and $h_a(w') = \epsilon$. Furthermore, since L_a is ϵ-free, $h_a(w) \ne \epsilon$ for each w in L_a. Let $q = \max\{q_a/a\}$ and consider h on $\tau'(L_2)$. If w' is a subword of a word of $\tau'(L_2)$, then w' is a subword of a word of the form $w_1 \ldots w_n$, with w_i in L_{a_i} for each i. Also, $w' = w_i'w_{i+1} \ldots w_{j-1}w_j'$, where $i \le j$, w_i' is a subword of w_i, and w_j' is a subword of w_j. Thus $|w_i'|\le q$ and $|w_j'|\le q$. Suppose $h(w) = \epsilon$. Then $w_{i+1}\ldots w_{j-1} = \epsilon$, so that $|w'|\le 2q$. If w' is a word of $\tau'(L_2)$, that is, $w' = w_1 \ldots w_n$, each w_i in L_{a_i}, and if $h(w) = \epsilon$; then $w_i = \epsilon$ for each i, so that $w = \epsilon$. Therefore h is restricted on $\tau'(L_2)$ and the proof of (a) is complete.

(b) Let L_2 be in $\mathcal{L}_2$, h a homomorphism restricted on L_2, and τ a substitution on $\Sigma^*_{h(L_2)}$ by languages of $\mathcal{L}_1$. Let c be a symbol not occurring in $\Sigma_{h(L)} \cup \tau h(L_2)$ and let τ' be the substitution on $\Sigma^*_{L_2}$ defined by $\tau'(a) = \{ch(a)\}$ for each a in Σ_{L_2}. Clearly $\tau'(L_2)$ is in $\mathrm{Sub}(\mathcal{F}_0, \mathcal{L}_2)$. Let τ'' be the substitution on $(\{c\} \cup \Sigma_{h(L)})^*$ defined by $\tau''(c) = \{c\}$ and $\tau''(b) = \tau(b)$ for b in $\Sigma_{h(L)}$. Then $\tau''[\tau'(L_2)]$ is in $\mathrm{Sub}[\mathcal{L}_1 \cup \mathcal{F}_0, \mathrm{Sub}(\mathcal{F}_0, \mathcal{L}_2)]$. Let h' be the homomorphism on $\Sigma^*_{\tau''\tau'(L_2)}$ defined by $h'(c) = \epsilon$ and $h'(b) = b$ for b in $\Sigma_{\tau''\tau'(L_2)} - \{c\}$. Obviously $\tau h(w) = h'\tau''\tau'(w)$ for each w in $\Sigma^*_{L_2}$.

To complete the proof it suffices to show that h' is restricted on $\tau''\tau'(L_2)$. Suppose there exists w in L_2 and w' in $\tau''\tau'(w)$ such that $h'(w') = \epsilon$. Then ϵ is in $\tau h(w)$. Since τ is a substitution by ϵ-free sets, $h(w) = \epsilon$. Since h is restricted on L_2, $w = \epsilon$. Hence $w' = \epsilon$. Suppose there exists $w \ne \epsilon$ in L_2 and a subword w' of some word in $\tau''\tau'(w)$ such that $h'(w') = \epsilon$. Then $w' = c^k$ for some $k>0$. Let $w = a_1 \ldots a_n$, each a_i in Σ_{L_2}. Then there exist positive

integers i and j, with $i < j$, and words $u_i, u_{i+1}, \ldots, u_{j-1}, u_j$ such that u_i and u_j are subwords of words in $\sigma\tau h(a_i)$ and $\sigma\tau h(a_j)$ resp, u_m is a word in $\sigma\tau h(a_m)$ for each m, $i < m < j$, and $w' = u_i u_{i+1} \cdots u_{j-1} u_j$. Thus $k \le |j-i| + 2$. Since $\mathcal{L}_1$ is ϵ-free, $h(a_m) = \epsilon$ for each m, $i < m < j$. Since h is restricted on L_2, there exists $q > 0$ such that $|v| < q$ if $h(v) = \epsilon$ and v is a subword of a word in L_2. Hence $|j-i| < q+2$, so that $|w'| = k < q+4$. Therefore h' is restricted on $\tau''\tau'(L_2)$.

Remark. The method of proof of Lemma 3.2 shows that both parts (a) and (b) hold if the ϵ-free condition on $\mathcal{L}_1$ is dropped and Hom_r is replaced throughout by Hom.

Lemma 3.3. Let $\mathcal{L}_2$ be a family of languages such that $\mathrm{Sub}(\mathcal{R}_o, \mathcal{L}_2) \subseteq \mathcal{L}_2$ and let $\mathcal{L}_3$ be a (full) AFL. Given an ϵ-free (arbitrary) symmetric family $\mathcal{L}$ containing $\mathcal{R}_o$ such that $\mathrm{Sub}(\mathcal{L}, \mathcal{L}_2) \subseteq \mathcal{L}_3$, then $\mathrm{Sub}(\mathrm{AFL}(\mathcal{L}), \mathcal{L}_2) \subseteq \mathcal{L}_3$ ($\mathrm{Sub}(\mathrm{AFL}_f(\mathcal{L}), \mathcal{L}_2) \subseteq \mathcal{L}_3$).

Proof. Let Δ be the collection of all ϵ-free (or arbitrary) symmetric families $\mathcal{L}'$ containing $\mathcal{R}_o$ and such that $\mathrm{Sub}(\mathcal{L}', \mathcal{L}_2) \subseteq \mathcal{L}_3$. Let $\{\mathcal{L}^{(n)}\}_{n \ge o}$ be defined with respect to $\mathcal{L}$ as in the corollary to Proposition 1.2 (the corollary to Proposition 1.3). Then $\mathcal{L}^{(o)} = \mathcal{L}$.

We shall show that $\mathcal{L}'$ in Δ implies $\mathrm{Sub}(\mathcal{R}_o, \mathcal{L}')$ (or $\mathrm{Sub}(\mathcal{R}, \mathcal{L}')$), $\mathrm{Sub}(\mathcal{L}', \mathcal{R}_o)$, $\mathcal{L}' \wedge \mathcal{R}$, and $\mathrm{Hom}_r(\mathcal{L}')$[15] are all in Δ. By induction on n it will then follow that each $\mathcal{L}^{(n)}$ is in Δ. Clearly each of the above sets is symmetric. Also,

---

[15] $\mathrm{Hom}_r(\mathcal{L}')$ is omitted if $\mathcal{L}$ is not ϵ-free.

$$\mathrm{Sub}[\mathrm{Sub}(\mathcal{R}_o, \mathcal{L}'),\mathcal{L}_2] = \mathrm{Sub}[\mathcal{R}_o, \mathrm{Sub}(\mathcal{L}',\mathcal{L}_2)] \subseteq \mathrm{Sub}(\mathcal{R}_o, \mathcal{L}_3) \subseteq \mathcal{L}_3$$

(or $\mathrm{Sub}[\mathrm{Sub}(\mathcal{R}, \mathcal{L}'), \mathcal{L}_2] = \mathrm{Sub}[\mathcal{R}, \mathrm{Sub}(\mathcal{L}', \mathcal{L}_2)] \subseteq \mathrm{Sub}(\mathcal{R}, \mathcal{L}_3) \subseteq \mathcal{L}_3$)

and $\mathrm{Sub}[\mathrm{Sub}(\mathcal{L}', \mathcal{R}_o),\mathcal{L}_2] = \mathrm{Sub}[\mathcal{L}', \mathrm{Sub}(\mathcal{R}_o,\mathcal{L}_2)] \subseteq \mathrm{Sub}(\mathcal{L}', \mathcal{L}_2) \subseteq \mathcal{L}_3$.

Clearly $\mathrm{Sub}(\mathcal{L}' \wedge \mathcal{R}, \mathcal{L}_2) \subseteq \mathrm{Sub}(\mathcal{L}' \wedge \mathcal{R}, \mathcal{L}_2 \wedge \mathcal{R}.)$ Since $\mathcal{R}_o \subseteq \mathcal{L}'$, $\mathcal{F}_o \subseteq \mathcal{L}'$. Thus, by Lemma 3.1,

$$\begin{aligned}\mathrm{Sub}(\mathcal{L}' \wedge \mathcal{R}, \mathcal{L}_2) &\subseteq \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}', \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_2)) \wedge \mathcal{R}]\\ &\subseteq \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}', \mathcal{L}_2)\wedge \mathcal{R}]\\ &\subseteq \mathrm{Hom}_r[\mathcal{L}_3 \wedge \mathcal{R}]\\ &\subseteq \mathcal{L}_3\end{aligned}$$

(or $\mathrm{Sub}(\mathcal{L}' \wedge \mathcal{R}, \mathcal{L}_2) \subseteq \mathrm{Hom}(\mathcal{L}_3 \wedge \mathcal{R}) \subseteq \mathcal{L}_3$).

By Lemma 3.2 (a),

$$\begin{aligned}\mathrm{Sub}[\mathrm{Hom}_r(\mathcal{L}'), \mathcal{L}_2) &\subseteq \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}', \mathcal{L}_2)]\\ &\subseteq \mathrm{Hom}_r(\mathcal{L}_3)\\ &\subseteq \mathcal{L}_3.\end{aligned}$$

It now follows that the sequence of families $\{\mathcal{L}^{(n)}\}_{n\geq 0}$ defined for $\mathcal{L}$ as in the corollary to Proposition 1.2 (or the corollary to Proposition 1.3) is in Δ. To complete the proof of the lemma it suffices to show that $\bigcup_{n\geq 0}\mathcal{L}^{(n)}$ is in Δ. Clearly $\bigcup_{n\geq 0}\mathcal{L}^{(n)}$ is ϵ-free if $\mathcal{L}$ is ϵ-free, contains $\mathcal{R}_o$, and is symmetric. Suppose τ is a substitution of L_2 in $\mathcal{L}_2$ by languages of $\bigcup_{n\geq 0}\mathcal{L}^{(n)}$. Then $\tau(L_2)$ involves only a finite number of languages of $\bigcup\mathcal{L}^{(n)}$. Since $\mathcal{L}^{(n)}$ is increasing in n, there exists m such that τ is a substitution by languages of $\mathcal{L}^{(m)}$. Hence

$$\mathrm{Sub}(\bigcup_{n\geq 0} \mathcal{L}^{(n)},\mathcal{L}_2) \subseteq \bigcup_{n\geq 0} \mathrm{Sub}(\mathcal{L}^{(n)}, \mathcal{L}_2) \subseteq \mathcal{L}_3.$$

Therefore $\bigcup_{n\geq 0}\mathcal{L}^{(n)}$ is in Δ and the proof is complete.

Next we have the analogue to Lemma 3.3 for the second variable in $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$.

Lemma 3.4. Let $\mathcal{L}_1$ be a family of languages containing $\mathcal{R}_o$ and such that $\mathrm{Sub}(\mathcal{L}_1, \mathcal{R}_o) \subseteq \mathcal{L}_1$. Let $\mathcal{L}_3$ be an AFL, with $\mathcal{L}_1 \subseteq \mathcal{L}_3$, and let $\mathcal{L}$ be a family of languages such that $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}) \subseteq \mathcal{L}_3$. If $\mathcal{L}_1$ is ϵ-free (or $\mathcal{L}_3$ is a full AFL), then $\mathrm{Sub}[\mathcal{L}_1, \mathrm{AFL}(\mathcal{L})] \subseteq \mathcal{L}_3$ (or $\mathrm{Sub}(\mathcal{L}_1, \mathrm{AFL}_f(\mathcal{L})) \subseteq \mathcal{L}_3$).

Proof. Let Δ be the collection of all families $\mathcal{L}'$ such that $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}') \subseteq \mathcal{L}_3$. Let $\{\mathcal{L}^{(n)}\}_{n\geq o}$ be the sequence of families defined for $\mathcal{L}$ in the corollary to Proposition 1.2 (or the corollary to Proposition 1.3). Since $\mathcal{L}^{(o)} = \mathcal{L} \cup \mathcal{R}_o$ and $\mathrm{Sub}(\mathcal{L}_1, \mathcal{R}_o) \subseteq \mathcal{L}_1 \subseteq \mathcal{L}_3$,

$$\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}^{(o)}) = \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}) \cup \mathrm{Sub}(\mathcal{L}_1, \mathcal{R}_o) \subseteq \mathcal{L}_3.$$

Therefore $\mathcal{L}^{(o)}$ is in Δ. We shall show that $\mathcal{L}'$ in Δ implies $\mathrm{Sub}(\mathcal{R}_o, \mathcal{L}')$ (or $\mathrm{Sub}(\mathcal{R}, \mathcal{L}')$), $\mathrm{Sub}(\mathcal{L}', \mathcal{R}_o)$, $\mathcal{L}' \wedge \mathcal{R}$, and $\mathrm{Hom}_r(\mathcal{L}')$ [16] are all in Δ.

Clearly

$$\begin{aligned} \mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{R}_o, \mathcal{L}')) &= \mathrm{Sub}(\mathrm{Sub}(\mathcal{L}_1, \mathcal{R}_o), \mathcal{L}') \\ &\subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}') \subseteq \mathcal{L}_3 \end{aligned}$$

$$\begin{aligned} (\text{or} \quad \mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}(\mathcal{R}, \mathcal{L}')) &= \mathrm{Sub}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{R}), \mathcal{L}'] \\ &= \mathrm{Sub}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{R}_o) \cup \{\epsilon\}, \mathcal{L}'] \\ &\subseteq \mathrm{Sub}[\mathcal{L}_1 \cup \{\epsilon\}, \mathcal{L}'] \\ &\subseteq \mathrm{HomSub}(\mathcal{L}_1, \mathcal{L}') \\ &\subseteq \mathcal{L}_3) \end{aligned}$$

$$\begin{aligned} \text{and} \quad \mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}(\mathcal{L}', \mathcal{R}_o)] &\subseteq \mathrm{Sub}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}'), \mathcal{R}_o], \text{ by (a) of Proposition 1.1,} \\ &\subseteq \mathrm{Sub}(\mathcal{L}_3, \mathcal{R}_o) \\ &\subseteq \mathcal{L}_3. \end{aligned}$$

[16] $\mathrm{Hom}_r(\mathcal{L}')$ is omitted if $\mathcal{L}_3$ is a full AFL.

Now $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}'\wedge \mathcal{R}) \subseteq \mathrm{Sub}(\mathcal{L}_1\wedge \mathcal{R}, \mathcal{L}'\wedge \mathcal{R})$

$$\subseteq \mathrm{Hom}_r(\mathrm{Sub}[\mathcal{L}_1\cup\mathcal{F}_o, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}')]\wedge\mathcal{R}), \text{ by Lemma 3.1,}$$

$$= \mathrm{Hom}_r(\mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}')] \wedge \mathcal{R})$$

$$= \mathrm{Hom}_r(\mathrm{Sub}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{F}_o), \mathcal{L}'] \wedge \mathcal{R})$$

$$\subseteq \mathrm{Hom}_r(\mathrm{Sub}[\mathcal{L}_1, \mathcal{L}'] \wedge \mathcal{R})$$

$$\subseteq \mathrm{Hom}_r(\mathcal{L}_3\wedge \mathcal{R}) \subseteq \mathcal{L}_3$$

(or $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}'\wedge \mathcal{R}) \subseteq \mathrm{Hom}(\mathcal{L}_3\wedge \mathcal{R}) \subseteq \mathcal{L}_3$).

By Lemma 3.2(b),

$$\mathrm{Sub}(\mathcal{L}_1, \mathrm{Hom}_r(\mathcal{L}')) \subseteq \mathrm{Hom}_r(\mathrm{Sub}[\mathcal{L}_1\cup \mathcal{F}_o, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}')])$$

$$\subseteq \mathrm{Hom}_r(\mathrm{Sub}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{F}_o), \mathcal{L}'])$$

$$\subseteq \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}')]$$

$$\subseteq \mathcal{L}_3.$$

From the above, it follows that $\mathcal{L}^{(n)}$ is in Δ for each $n\geq 0$. Then

$$\mathrm{Sub}(\mathcal{L}_1, \mathrm{AFL}(\mathcal{L})) = \mathrm{Sub}(\mathcal{L}_1, \bigcup_{n\geq o} \mathcal{L}^{(n)})$$

$$\subseteq \bigcup_{n\geq o} \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}^{(n)})$$

$$\subseteq \mathcal{L}_3$$

(or, similarly, $\mathrm{Sub}(\mathcal{L}_1, \mathrm{AFL}_f(\mathcal{L})) \subseteq \mathcal{L}_3$) and the proof is complete.

<u>Lemma 3.5</u>. Let $\mathcal{L}_1$ be an ϵ-free (arbitrary) family of languages containing $\mathcal{F}_o$ such that $\mathrm{Sub}(\mathcal{L}_1, \mathcal{F}_o) \subseteq \mathcal{L}_1$, and let $\mathcal{L}_2$ be a nontrivial family of languages. Then $\mathcal{L}_1 \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$ (or $\mathcal{L}_1 \subseteq \mathrm{AFL}_f[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$).

<u>Proof</u>. Let L_2 be a language in $\mathcal{L}_2$ containing a word of length $k\geq 1$ and let L_1 be a language in $\mathcal{L}_1$. [L_2 exists since $\mathcal{L}_2$ is nontrivial.] Let c be a new symbol. Since $\mathcal{F}_o \subseteq \mathcal{L}_1$, $\{c\}$ is in $\mathcal{L}_1$. Since $\mathrm{Sub}(\mathcal{L}_1, \mathcal{F}_o) \subseteq \mathcal{L}_1$, L_1c, hence $L_1c \cup \{c\}$, is in $\mathcal{L}_1$. Let τ be the substitution on $\Sigma_{L_2}^*$ defined by $\tau(a) = L_1c \cup \{c\}$

for each a in Σ_{L_2}. Then $\tau(L_2)$ is in $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$. Thus $\tau(L_2) \cap \Sigma_{L_1}^{+} c^k$ and $\tau(L_2) \cap \Sigma_{L_1}^{*} c^k$ are in $\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$. Then $L_1 c^k = \tau(L_2) \cap \Sigma_{L_1}^{+} c^k$ ($L_1 c^k = \tau(L_2) \cap \Sigma_{L_1}^{*} c^k$ or $L_1 c^k = \tau(L_2) \cap \Sigma_{L_1}^{*} c^k$). Thus $L_1 c^k$ is in $\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$ ($\mathrm{AFL}_f[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$). Let h be the homomorphism on $(\Sigma_{L_1} \cup \{c\})^*$ defined by $h(c) = \epsilon$ and $h(b) = b$ for b in Σ_{L_1}. Then h is restricted (arbitrary homomorphism) on $L_1 c^k$ and $L_1 = h(L_1 c^k)$ is in $\mathrm{Hom}_r[\mathrm{AFL}(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2))] \subseteq \mathrm{AFL}\,(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2))$ (or L_1 is in $\mathrm{Hom}[\mathrm{AFL}_f(\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] \subseteq \mathrm{AFL}_f[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$).

We are now ready for the main result of the section.

<u>Theorem 3.1</u>. Let $\mathcal{L}_1$ be an ϵ-free (or arbitrary) symmetric family of languages containing $\mathcal{R}_0$ and let $\mathcal{L}_2$ be a nontrivial family of languages. If either $\mathrm{Sub}(\mathcal{R}_0, \mathcal{L}_2) \subseteq \mathcal{L}_2$ or $\mathrm{Sub}(\mathcal{L}_1, \mathcal{R}_0) \subseteq \mathcal{L}_1$, then

$$\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] = \mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1), \mathrm{AFL}(\mathcal{L}_2)]$$

(or $\quad \mathrm{AFL}_f[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] = \mathrm{Sub}[\mathrm{AFL}_f(\mathcal{L}_1), \mathrm{AFL}_f(\mathcal{L}_2)]$.)

<u>Proof</u>. We only consider the case when $\mathcal{L}_1$ is ϵ-free, since the other case can be treated similarly.

By Theorem 2.1, $\mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1), \mathrm{AFL}(\mathcal{L}_2)]$ is an AFL. Since this AFL contains $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$, it follows that

$$\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] \subseteq \mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1), \mathrm{AFL}(\mathcal{L}_2)].$$

Consider the reverse inclusion. First assume that $\mathrm{Sub}(\mathcal{R}_0, \mathcal{L}_2) \subseteq \mathcal{L}_2$. Since $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$, by Lemma 3.3 we obtain

(*) $\quad \mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1), \mathcal{L}_2] \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$.

Since $\mathcal{L}_2$ is a nontrivial family,

$\mathrm{AFL}(\mathcal{L}_1) \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathrm{AFL}(\mathcal{L}_1), \mathcal{L}_2)]$, by Lemma 3.5,

(**) $\qquad \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$, by (*).

By (**), (*), and Lemma 3.4,

$$\mathrm{Sub}(\mathrm{AFL}(\mathcal{L}_1),\ \mathrm{AFL}(\mathcal{L}_2)] \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1,\ \mathcal{L}_2)].$$

Next assume that $\mathrm{Sub}(\mathcal{L}_1, \mathcal{R}_0) \subseteq \mathcal{L}_1$. Since $\mathcal{L}_2$ is a nontrivial family, $\mathcal{L}_1 \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$ by Lemma 3.5. Since $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$,

$$\mathrm{Sub}[\mathcal{L}_1,\ \mathrm{AFL}(\mathcal{L}_2)] \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1,\ \mathcal{L}_2)]$$

by Lemma 3.4. Then, by Lemma 3.3,

$$\mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1),\ \mathrm{AFL}(\mathcal{L}_2)] \subseteq \mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1,\ \mathcal{L}_2)].$$

Remarks. (1) The proof of Theorem 3.1 shows that the inclusion

$$\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1,\ \mathcal{L}_2)] \subseteq \mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1),\ \mathrm{AFL}(\mathcal{L}_2)]$$

(or $\mathrm{AFL}_f[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] \subseteq \mathrm{Sub}[\mathrm{AFL}_f(\mathcal{L}_1), \mathrm{AFL}_f(\mathcal{L}_2)]$)

is valid with no hypotheses on $\mathcal{L}_1$ or $\mathcal{L}_2$.

(2) The reverse inclusion in remark (1), thus Theorem 3.1, is not valid without some hypotheses. For example, if $\mathcal{L}_1$ is the trivial family consisting of just $\{\epsilon\}$, then $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_1$ and $\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] = \mathcal{R} = \mathrm{AFL}_f[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$ need not contain $\mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1), \mathrm{AFL}(\mathcal{L}_2)] = \mathrm{AFL}_f(\mathcal{L}_2) = \mathrm{Sub}[\mathrm{AFL}_f(\mathcal{L}_1), \mathrm{AFL}_f(\mathcal{L}_2)]$. Similarly, if $\mathcal{L}_2$ is the trivial family, then $\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] = \mathcal{R} = \mathrm{AFL}_f[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)]$ need not contain $\mathrm{Sub}(\mathrm{AFL}[\mathcal{L}_1), \mathcal{R})$ which contains $\mathrm{AFL}(\mathcal{L}_1)$. If $\mathcal{L}_1 = a^+$ and $\mathcal{L}_2$ is the AFL generated by $\{a^nb^n/n\geq 1\}$, then $\mathrm{Sub}(\mathcal{R}_0, \mathcal{L}_2) = \mathcal{L}_2$, $\mathrm{AFL}[\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)] \subseteq \mathrm{AFL}(\mathcal{L}_1) = \mathcal{R}_0$, and $\mathrm{Sub}[\mathrm{AFL}(\mathcal{L}_1), \mathrm{AFL}(\mathcal{L}_2)] = \mathrm{Sub}(\mathcal{R}_0, \mathcal{L}_2) = \mathcal{L}_2$.

Section 4. Iterated substitution

We are interested in families of languages closed with respect to substitution into another family, or closed with respect to substitution by languages

of another family. If a family lacks the desired closure property, then we apply the substitutions in question and enlarge the original family by adding the languages obtained by these substitutions. Iterating this procedure leads to a family containing the original one and having the desired closure property.

This section contains the definitions and some results about iterated substitution. In particular, we show that iterated substitution applied to an AFL yields an AFL. We present a condition that guarantees the substitution closure of a family to be an AFL even if the original family is not an AFL.

Definition. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be families of languages. Then $\mathcal{L}_2$ is closed under $\mathcal{L}_1$ substitution if $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathcal{L}_2$. The closure of $\mathcal{L}_2$ under $\mathcal{L}_1$ substitution is the smallest family containing $\mathcal{L}_2$ and closed under $\mathcal{L}_1$ substitution.

Since the intersection of all families containing $\mathcal{L}_2$ and closed under $\mathcal{L}_1$ substitution (the family of all sets $L \subseteq \Sigma_1^* \subseteq \Sigma^*$ is one such family) is also such a family, the closure exists. In Lemma 4.1, we shall show how to calculate it.

Notation. Given $\mathcal{L}_1$ and $\mathcal{L}_2$, let $\mathrm{Sub}^0(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_2$ and, by induction, $\mathrm{Sub}^{k+1}(\mathcal{L}_1, \mathcal{L}_2) = \mathrm{Sub}[\mathcal{L}_1, \bigcup_{i=0}^{k} \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)]$ for each $k \geq 0$. Let $\mathrm{Sub}^\infty(\mathcal{L}_1, \mathcal{L}_2) = \bigcup_{i=0}^{\infty} \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)$.

Note that for every $\mathcal{L}_1$, $\mathcal{L}_2$, and k,

$$\begin{aligned}\mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}^k(\mathcal{L}_1, \mathcal{L}_2)] &\subseteq \bigcup_{i=0}^{k} \mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)] \\ &= \mathrm{Sub}[\mathcal{L}_1, \bigcup_{i=0}^{k} \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)] \\ &\subseteq \mathrm{Sub}^{k+1}(\mathcal{L}_1, \mathcal{L}_2).\end{aligned}$$

Also, for each $k \geq 1$,

$$\begin{aligned}\mathrm{Sub}^k(\mathcal{L}_1, \mathcal{L}_2) &= \mathrm{Sub}[\mathcal{L}_1, \bigcup_{o}^{k-1} \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)] \\ &\subseteq \mathrm{Sub}[\mathcal{L}_1, \bigcup_{o}^{k} \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)] \\ &= \mathrm{Sub}^{k+1}(\mathcal{L}_1, \mathcal{L}_2).\end{aligned}$$

It is not necessarily true that $\mathcal{L}_2 = \mathrm{Sub}^o(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathrm{Sub}^1(\mathcal{L}_1, \mathcal{L}_2)$. [For example, let $\mathcal{L}_1 = \{\{ab\}\}$ and $\mathcal{L}_2 = \{\{a\}\}$. Thus $\mathrm{Sub}^{(1)}(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_1$, which does not contain $\mathcal{L}_2 = \mathrm{Sub}^o(\mathcal{L}_1, \mathcal{L}_2)$.]

Suppose $\mathcal{L}_1$ contains $\{a\}$ for each a in Σ. Then $\mathcal{L}_2 \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2)$ for every family $\mathcal{L}_2$. Thus, for every $k \geq 0$,

$$\begin{aligned}\mathrm{Sub}^{k+1}(\mathcal{L}_1, \mathcal{L}_2) &= \mathrm{Sub}[\mathcal{L}_1, \bigcup_{i=o}^{k} \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)] \\ &= \mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}^k(\mathcal{L}_1, \mathcal{L}_2)].\end{aligned}$$

<u>Lemma 4.1.</u> The family $\mathrm{Sub}^\infty(\mathcal{L}_1, \mathcal{L}_2)$ is the closure of $\mathcal{L}_2$ under $\mathcal{L}_1$ substitution.

<u>Proof.</u> Let $\mathcal{L}$ be the closure of $\mathcal{L}_2$ with respect to $\mathcal{L}_1$ substitution. Since $\mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}^k(\mathcal{L}_1,\mathcal{L}_2)] \subseteq \mathrm{Sub}^{k+1}(\mathcal{L}_1, \mathcal{L}_2)$,

$$\begin{aligned}\mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}^\infty(\mathcal{L}_1, \mathcal{L}_2)] &= \mathrm{Sub}[\mathcal{L}_1, \bigcup_{o}^{\infty} \mathrm{Sub}^k(\mathcal{L}_1, \mathcal{L}_2)] \\ &= \bigcup_{o}^{\infty} \mathrm{Sub}[\mathcal{L}_1, \mathrm{Sub}^k(\mathcal{L}_1, \mathcal{L}_2)] \\ &\subseteq \bigcup_{o}^{\infty} \mathrm{Sub}^{k+1}(\mathcal{L}_1, \mathcal{L}_2) \\ &\subseteq \mathrm{Sub}^\infty(\mathcal{L}_1, \mathcal{L}_2).\end{aligned}$$

Thus $\mathrm{Sub}^\infty(\mathcal{L}_1, \mathcal{L}_2)$ is closed under $\mathcal{L}_1$ substitution, so that $\mathcal{L} \subseteq \mathrm{Sub}^\infty(\mathcal{L}_1, \mathcal{L}_2)$.

To see the reverse containment, it suffices to show that $\mathrm{Sub}^k(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathcal{L}$ for each $k \geq 0$. By definition, $\mathrm{Sub}^o(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_2 \subseteq \mathcal{L}$. Continuing by induction, suppose $\mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathcal{L}$ for $0 \leq i < k$, $k \geq 1$. Then

$$\mathrm{Sub}^k(\mathcal{L}_1, \mathcal{L}_2) = \mathrm{Sub}[\mathcal{L}_1, \bigcup_{o}^{k-1} \mathrm{Sub}^i(\mathcal{L}_1, \mathcal{L}_2)] \subseteq \mathrm{Sub}(\mathcal{L}_1, \mathcal{L}) \subseteq \mathcal{L}.$$

Thus the induction is extended and the proof is complete.

Definition. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be families of languages. A family $\mathcal{L}_1$ is closed under substitution in $\mathcal{L}_2$ if $\mathrm{Sub}(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathcal{L}_1$. The closure of $\mathcal{L}_1$ under substitution in $\mathcal{L}_2$ is the smallest family containing $\mathcal{L}_1$ and closed under substitution in $\mathcal{L}_2$.

Again, it is easily seen that the closure exists.

Notation. Given $\mathcal{L}_1$ and $\mathcal{L}_2$, let $\mathrm{Sub}_o(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_1$ and, by induction, $\mathrm{Sub}_{k+1}(\mathcal{L}_1, \mathcal{L}_2) = \mathrm{Sub}[\bigcup_{i=o}^{k} \mathrm{Sub}_i(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_2]$ for each $k \geq 0$. Let $\mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2) = \bigcup_{i=o}^{\infty} \mathrm{Sub}_i(\mathcal{L}_1, \mathcal{L}_2)$.

Again, $\mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathrm{Sub}_{k+1}(\mathcal{L}_1, \mathcal{L}_2)$ for $k \geq 1$. If $\mathcal{L}_2$ contains $\{a\}$ for some a in Σ, then $\mathrm{Sub}_o(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_1 \subseteq \mathrm{Sub}_1(\mathcal{L}_1, \mathcal{L}_2)$.

Lemma 4.2. The family $\mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2)$ is the closure of $\mathcal{L}_1$ under substitution in $\mathcal{L}_2$.

Proof. Let $\mathcal{L}$ be the closure of $\mathcal{L}_1$ under substitution in $\mathcal{L}_2$. As in the proof of Lemma 4.1, it is easily seen that $\mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathcal{L}$ for each $k \geq 0$. Hence $\mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathcal{L}$. To see the reverse containment, we have

$$\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_2] = \mathrm{Sub}[\bigcup_{o}^{\infty} \mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_2].$$

Let L_2 be in $\mathcal{L}_2$ and τ a substitution of L_2 by languages of $\bigcup_{o}^{\infty} \mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2)$. Since Σ_{L_2} is finite, there exists $m \geq 0$ such that $\tau(a)$ is in $\bigcup_{o}^{m} \mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2)$ for each a in Σ_2. Therefore $\tau(L_2)$ is in $\mathrm{Sub}[\bigcup_{o}^{m} \mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_2] = \mathrm{Sub}_{m+1}(\mathcal{L}_1, \mathcal{L}_2) \subseteq \mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2)$. Thus

$$\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_2] \subseteq \mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2),$$

so that $\mathcal{L} \subseteq \mathrm{Sub}_\infty(\mathcal{L}_1, \mathcal{L}_2)$.

Definition. A family $\mathcal{L}$ is substitution closed if $\mathrm{Sub}(\mathcal{L},\mathcal{L}) \subseteq \mathcal{L}$. The substitution closure of $\mathcal{L}$ is the smallest substitution closed family $\mathcal{L}_\infty$ containing $\mathcal{L}$.

Clearly $\mathcal{L}_\infty$ exists.

Theorem 4.1. (a) For each family $\mathcal{L}$, $\mathcal{L}_\infty = \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L})$.

(b) If $\mathcal{L}$ is a symmetric family such that $\mathcal{L} \subseteq \mathrm{Sub}(\mathcal{L},\mathcal{L})$, then $\mathcal{L}_\infty = \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})$.

Proof. (a) Since $\mathcal{L}_\infty$ is substitution closed and contains $\mathcal{L}$, it is closed under substitution in $\mathcal{L}$. By Lemma 4.2, $\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}) \subseteq \mathcal{L}_\infty$.

To complete the proof of (a), it suffices to show that $\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L})$ is substitution closed. [For this will imply that $\mathcal{L}_\infty \subseteq \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L})$.] Since $\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L})$ is the closure of $\mathcal{L}$ under substitution in $\mathcal{L}$,

$$\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathrm{Sub}_0(\mathcal{L},\mathcal{L})] = \mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathcal{L}] \subseteq \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}).$$

Continuing by induction, assume $n>0$ and that $\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathrm{Sub}_j(\mathcal{L},\mathcal{L})] \subseteq \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L})$ for all $0 \leq j < n$. Then for $\mathcal{L}' = \bigcup_{j=0}^{n-1} \mathrm{Sub}_j(\mathcal{L},\mathcal{L})$,

$$\begin{aligned}\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathcal{L}'] &= \bigcup_{j=0}^{n-1} \mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathrm{Sub}_j(\mathcal{L},\mathcal{L})] \\ &\subseteq \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}).\end{aligned}$$

Furthermore, $\mathrm{Sub}_n(\mathcal{L},\mathcal{L}) = \mathrm{Sub}(\mathcal{L}',\mathcal{L})$. Thus

$$\begin{aligned}\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathrm{Sub}_n(\mathcal{L},\mathcal{L})] &= \mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathrm{Sub}(\mathcal{L}',\mathcal{L})] \\ &\subseteq \mathrm{Sub}(\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}),\mathcal{L}'], \mathcal{L}), \\ &\quad \text{by (a) of Proposition 1.1,} \\ &\subseteq \mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}), \mathcal{L}] \\ &\subseteq \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}).\end{aligned}$$

Hence the induction is extended. Therefore

$$\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}_k(\mathcal{L},\mathcal{L})] \subseteq \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L})$$

for all $k \geq 0$, so that

$$\begin{aligned}\mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L})] &= \bigcup_{k=0}^{\infty} \mathrm{Sub}[\mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}_k(\mathcal{L},\mathcal{L})] \\ &\subseteq \mathrm{Sub}_\infty(\mathcal{L},\mathcal{L}).\end{aligned}$$

(b) Since $\mathcal{L}_\infty$ is substitution closed and contains $\mathcal{L}$, it is closed under $\mathcal{L}$ substitution. By Lemma 4.1, $\mathrm{Sub}^\infty(\mathcal{L},\mathcal{L}) \subseteq \mathcal{L}_\infty$.

To complete the proof of (b), it suffices to show that $\mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})$ is substitution closed. By hypothesis, $\mathrm{Sub}^0(\mathcal{L},\mathcal{L}) = \mathcal{L} \subseteq \mathrm{Sub}^1(\mathcal{L},\mathcal{L})$. Thus $\mathrm{Sub}^k(\mathcal{L},\mathcal{L}) \subseteq \mathrm{Sub}^{k+1}(\mathcal{L},\mathcal{L})$ for all $k \geq 0$. Hence $\mathrm{Sub}^{k+1}(\mathcal{L},\mathcal{L}) = \mathrm{Sub}[\mathcal{L},\ \mathrm{Sub}^k(\mathcal{L},\mathcal{L})]$ for all $k \geq 0$. Obviously $\mathrm{Sub}^{k+1}(\mathcal{L},\mathcal{L})$ is symmetric for all $k \geq 0$. By Lemma 4.1,

$$\begin{aligned}\mathrm{Sub}[\mathcal{L},\ \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})] &= \mathrm{Sub}[\mathrm{Sub}^0(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})] \\ &\subseteq \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L}).\end{aligned}$$

Continuing by induction, assume that

$$\mathrm{Sub}[\mathrm{Sub}^k(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})] \subseteq \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})$$

for $0 \leq k < n$, $n \geq 1$. Then

$$\begin{aligned}\mathrm{Sub}[\mathrm{Sub}^n(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})] &= \mathrm{Sub}[\mathrm{Sub}(\mathcal{L},\ \mathrm{Sub}^{n-1}(\mathcal{L},\mathcal{L})), \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})] \\ &= \mathrm{Sub}[\mathcal{L},\ \mathrm{Sub}(\mathrm{Sub}^{n-1}(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L}))], \\ &\quad \text{by Proposition 1.1,} \\ &\subseteq \mathrm{Sub}[\mathcal{L},\ \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})] \\ &\subseteq \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L}).\end{aligned}$$

Therefore the induction is extended, so that

$$\mathrm{Sub}[\mathrm{Sub}^k(\mathcal{L},\mathcal{L}),\ \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})] \subseteq \mathrm{Sub}^\infty(\mathcal{L},\mathcal{L}).$$

Now let L be in $\mathrm{Sub}^\infty(\mathcal{L},\mathcal{L})$ and τ a substitution on Σ_L by languages in

$\mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L}) = \bigcup_{k=0}^{\infty} \mathrm{Sub}^{k}(\mathcal{L},\mathcal{L})$. There exists $m \geq 0$ such that $\tau(a)$ is in $\bigcup_{k=0}^{m} \mathrm{Sub}^{k}(\mathcal{L},\mathcal{L}) = \mathrm{Sub}^{m}(\mathcal{L},\mathcal{L})$ for each a in Σ_L. Thus $\tau(L)$ is in $\mathrm{Sub}[\mathrm{Sub}^{m}(\mathcal{L},\mathcal{L}), \mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L})]$. Hence

$$\begin{aligned}\mathrm{Sub}[\mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L}), \mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L})] &\subseteq \bigcup_{k=0}^{\infty} \mathrm{Sub}[\mathrm{Sub}^{k}(\mathcal{L},\mathcal{L}), \mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L})]\\ &\subseteq \mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L}),\end{aligned}$$

completing the proof.

We do not know to what extent the hypotheses in the theorems of Section 4 can be weakened. In particular, we do not know if the hypotheses of Theorem 4.1 can be weakened. In general, (b) of Theorem 4.1 is not valid without some hypotheses on $\mathcal{L}$, i.e., $\mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L})$ is not always substitution closed. For example, let $\mathcal{L} = \{\{ab\}/a,b \text{ in } \Sigma\}$. Then $\mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L})$ consists of all words of length 2^n for all $n \geq 1$. However, the substitution closure of $\mathcal{L}$ contains $\{a^6\}$ for each a in Σ.

The rest of this section is concerned with relations between the various substitution closures and AFL.

<u>Theorem 4.2</u>. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are AFL, then so are $\mathrm{Sub}^{\infty}(\mathcal{L}_1, \mathcal{L}_2)$ and $\mathrm{Sub}_{\infty}(\mathcal{L}_1, \mathcal{L}_2)$. If, in addition, $\mathcal{L}_1$ is full, then so are $\mathrm{Sub}^{\infty}(\mathcal{L}_1, \mathcal{L}_2)$ and $\mathrm{Sub}_{\infty}(\mathcal{L}_1,\mathcal{L}_2)$.

<u>Proof</u>. Since $\mathcal{L}_1$ contains $\{a\}$ for each a in Σ, $\mathrm{Sub}^{k}(\mathcal{L}_1, \mathcal{L}_2) = \mathrm{Sub}(\mathcal{L}_1, \mathrm{Sub}^{k-1}(\mathcal{L}_1, \mathcal{L}_2))$. Similarly, $\mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2) = \mathrm{Sub}(\mathrm{Sub}_{k-1}(\mathcal{L}_1, \mathcal{L}_2), \mathcal{L}_2)$. By Theorem 2.1 and induction, $\mathrm{Sub}^{k}(\mathcal{L}_1, \mathcal{L}_2)$ and $\mathrm{Sub}_k(\mathcal{L}_1, \mathcal{L}_2)$ are AFL for all k (and are full if $\mathcal{L}_1$ is full, by Corollary 1 of Theorem 2.1). Therefore $\mathrm{Sub}^{\infty}(\mathcal{L}_1, \mathcal{L}_2)$ and $\mathrm{Sub}_{\infty}(\mathcal{L}_1, \mathcal{L}_2)$ are AFL (and full if $\mathcal{L}_1$ is full).

From Theorems 4.1 and 4.2 we get

<u>Corollary</u>. For each (full) AFL $\mathcal{L}$, $\mathcal{L}_{\infty} = \mathrm{Sub}^{\infty}(\mathcal{L},\mathcal{L}) = \mathrm{Sub}_{\infty}(\mathcal{L},\mathcal{L})$ and is a (full) AFL.

<u>Remarks</u>. (1) For AFL $\mathcal{L}_1$ and $\mathcal{L}_2$, $\mathrm{Sub}^{\infty}(\mathcal{L}_1, \mathcal{L}_2)$ and $\mathrm{Sub}_{\infty}(\mathcal{L}_1, \mathcal{L}_2)$ are not necessarily

the same. For let $\mathcal{L}_1 = R_0$ and $\mathcal{L}_2$ be the family of context-free lang[illegible]ages. Thus $\mathrm{Sub}^{\infty}(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_2$, $\mathrm{Sub}_{\infty}(\mathcal{L}_1, \mathcal{L}_2) = \mathcal{L}_1$, and $\mathcal{L}_1 \neq \mathcal{L}_2$.

(2) For well-known AFL there arise the general problems of "[illegible]ifying" $\mathrm{Sub}^{\infty}(\mathcal{L}_1, \mathcal{L}_2)$ and $\mathrm{Sub}_{\infty}(\mathcal{L}_1, \mathcal{L}_2)$.

Theorem 4.3. Let $\mathcal{L}$ be an ϵ-free (or arbitrary) symmetric family of languages containing R_0 and such that either $\mathrm{Sub}(R_0, \mathcal{L}) \subseteq \mathcal{L}$ or $\mathrm{Sub}(\mathcal{L}, R_0) \subseteq \mathcal{L}$. Then $\mathrm{AFL}(\mathcal{L}_{\infty}) = [\mathrm{AFL}(\mathcal{L})]_{\infty}$ (or $\mathrm{AFL}_f(\mathcal{L}_{\infty}) = [\mathrm{AFL}_f(\mathcal{L})]_{\infty}$).

Proof. We shall prove the theorem fo[illegible] the ϵ-free case, an analogous argument holding for the arbitrary $\mathcal{L}$ case.

Note that since $R_0 \subseteq \mathcal{L}$ and $R_0 \subseteq \mathrm{AFL}(\mathcal{L})$, $\mathcal{L} \subseteq \mathrm{Sub}(\mathcal{L},\mathcal{L})$ and $\mathrm{AFL}(\mathcal{L}) \subseteq \mathrm{Sub}[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})]$. Hence, for each $n > 0$,

$$(*) \quad \mathrm{Sub}_n(\mathcal{L},\mathcal{L}) = \mathrm{Sub}[\mathrm{Sub}_{n-1}(\mathcal{L},\mathcal{L}), \mathcal{L}]$$

and $$(**) \quad \mathrm{Sub}_n[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})] = \mathrm{Sub}(\mathrm{Sub}_{n-1}[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})], \mathrm{AFL}(\mathcal{L})).$$

We first show that for each $n \geq 0$, $\mathrm{AFL}[\mathrm{Sub}_n(\mathcal{L},\mathcal{L})] = \mathrm{Sub}_n[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})]$. For $n = 0$, $\mathrm{AFL}[\mathrm{Sub}_n(\mathcal{L},\mathcal{L})] = \mathrm{AFL}(\mathcal{L}) = \mathrm{Sub}_n[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})]$. Continuing by induction, suppose $n>0$ and that $\mathrm{AFL}[\mathrm{Sub}_{n-1}(\mathcal{L},\mathcal{L})] = \mathrm{Sub}_{n-1}[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})]$. Then

$$\begin{aligned}
\mathrm{AFL}[\mathrm{Sub}_n(\mathcal{L},\mathcal{L})] &= \mathrm{AFL}(\mathrm{Sub}[\mathrm{Sub}_{n-1}(\mathcal{L},\mathcal{L}), \mathcal{L}]), \text{ by } (*), \\
&= \mathrm{Sub}(\mathrm{AFL}[\mathrm{Sub}_{n-1}(\mathcal{L},\mathcal{L})], \mathrm{AFL}(\mathcal{L})], \text{ by Theorem 3.1,}^{(17)} \\
&= \mathrm{Sub}(\mathrm{Sub}_{n-1}[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})], \mathrm{AFL}(\mathcal{L})), \text{ by induction,} \\
&= \mathrm{Sub}_n[\mathrm{AFL}(\mathcal{L}), \mathrm{AFL}(\mathcal{L})], \text{ by } (**),
\end{aligned}$$

extending the induction.

---

(17) Since $\mathcal{L}$ is symmetric, $\mathrm{Sub}_{n-1}(\mathcal{L},\mathcal{L})$ is symmetric. If $\mathrm{Sub}(R_0, \mathcal{L}) \subseteq \mathcal{L}$, then the hypotheses of Theorem 3.1 are obviously satisfied. Suppose $\mathrm{Sub}(\mathcal{L}, R_0) \subseteq \mathcal{L}$. A simple induction, using Proposition 1.1, shows that $\mathrm{Sub}[\mathrm{Sub}_k(\mathcal{L},\mathcal{L}), R_0] \subseteq \mathrm{Sub}_k(\mathcal{L},\mathcal{L})$ for each $k \geq 0$. Thus, in this case also, the hypotheses of Theorem 3.1 are satisfied.

To complete the proof, we see that $AFL(\mathcal{L}_\infty) = AFL[Sub_\infty(\mathcal{L},\mathcal{L})] =$ $AFL[\bigcup_n Sub_n(\mathcal{L},\mathcal{L})] = \bigcup_n AFL[Sub_n(\mathcal{L},\mathcal{L}] = \bigcup_n Sub_n[AFL(\mathcal{L}), AFL(\mathcal{L})] = [AFL(\mathcal{L})]_\infty$.

Corollary. If $\mathcal{L}$ is a substitution closed AFL, then the full AFL generated by $\mathcal{L}$ is also substitution closed.

Proof. By Theorem 4.3, $[AFL_f(\mathcal{L})]_\infty = AFL_f(\mathcal{L}_\infty)$. Since $\mathcal{L}$ is substitution closed, $\mathcal{L}_\infty = \mathcal{L}$. Hence $[AFL_f(\mathcal{L})]_\infty = AFL_f(\mathcal{L})$, so that $AFL_f(\mathcal{L})$is substitution closed.

Our final theorem provides a criterion for $\mathcal{L}_\infty$ to be an AFL even if $\mathcal{L}$ is not.

Theorem 4.4. (a) Let $\mathcal{L}$ be a family of languages containing Σ_1^+ for every finite $\Sigma_1 \subseteq \Sigma$ and such that $Sub(\mathcal{F}_0, \mathcal{L}) \subseteq \mathcal{L}$. Then $\mathcal{L}_\infty$ is an AFL if and only if $Hom_r(\mathcal{L} \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$.

(b) Let $\mathcal{L}$ be a family of languages containing Σ_1^* for every finite $\Sigma_1 \subseteq \Sigma$ and such that $Sub(\mathcal{F}_0, \mathcal{L}) \subseteq \mathcal{L}$. Then $\mathcal{L}_\infty$ is a full AFL if and only if $Hom(\mathcal{L} \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$, or equivalently, if and only if $\mathcal{L} \wedge \mathcal{R} \subseteq \mathcal{L}_\infty$.

Proof. (a) Suppose $\mathcal{L}_\infty$ is an AFL. Then $Hom_r(\mathcal{L}_\infty \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$. Since $\mathcal{L} \subseteq \mathcal{L}_\infty$, $Hom_r(\mathcal{L} \wedge \mathcal{R}) \subseteq Hom_r(\mathcal{L}_\infty \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$.

To prove the sufficiency, assume $Hom_r(\mathcal{L} \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$. Let L be in $\mathcal{R}_0$. Then Σ_L^+ is in $\mathcal{L}$, so that $L = \Sigma_L^+ \cap L \subseteq Hom_r(\mathcal{L} \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$. Thus $\mathcal{R}_0 \subseteq \mathcal{L}_\infty$. Since $\mathcal{L}_\infty$ is substitution closed,

$$Sub(\mathcal{R}_0, \mathcal{L}_\infty) \subseteq Sub(\mathcal{L}_\infty, \mathcal{L}_\infty) \subseteq \mathcal{L}_\infty$$

and

$$Sub(\mathcal{L}_\infty, \mathcal{R}_0) \subseteq Sub(\mathcal{L}_\infty, \mathcal{L}_\infty) \subseteq \mathcal{L}_\infty.$$

Thus, to show that $\mathcal{L}_\infty$ is an AFL, it suffices to verify that $\mathcal{L}_\infty \wedge \mathcal{R} \subseteq \mathcal{L}_\infty$ and $Hom_r(\mathcal{L}_\infty) \subseteq \mathcal{L}_\infty$. Since $(\mathcal{L}_\infty \wedge \mathcal{R}) \cup Hom_r(\mathcal{L}_\infty) \subseteq Hom_r(\mathcal{L}_\infty \wedge \mathcal{R})$, it

suffices to show that $\mathrm{Hom}_r(\mathcal{L}_\infty \wedge R) \subseteq \mathcal{L}_\infty$.

We first show that $\mathrm{Hom}_r[\mathrm{Sub}_n(\mathcal{L},\mathcal{L}) \wedge R] \subseteq \mathcal{L}_\infty$ for each $n \geq 0$. For $n = 0$, $\mathrm{Hom}_r[\mathrm{Sub}_n(\mathcal{L},\mathcal{L}) \wedge R] = \mathrm{Hom}_r(\mathcal{L} \wedge R) \subseteq \mathcal{L}_\infty$, by hypothesis. Assume $n>0$ and $\mathrm{Hom}_r[\mathrm{Sub}_j(\mathcal{L},\mathcal{L}) \wedge R] \subseteq \mathcal{L}_\infty$ for all j, $0 \leq j < n$. Let $\mathcal{L}' = \bigcup_{j=0}^{n-1} \mathrm{Sub}_j(\mathcal{L},\mathcal{L})$. Then $\mathrm{Hom}_r(\mathcal{L}' \wedge R) \subseteq \mathcal{L}_\infty$ and, by definition, $\mathrm{Sub}_n(\mathcal{L},\mathcal{L}) = \mathrm{Sub}(\mathcal{L}',\mathcal{L})$. Therefore

$$\begin{aligned}
\mathrm{Hom}_r[\mathrm{Sub}_n(\mathcal{L},\mathcal{L}) \wedge R] &= \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{L}',\mathcal{L}) \wedge R] \\
&\subseteq \mathrm{Hom}_r(\mathrm{Sub}[(\mathcal{L}' \wedge R), \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}) \wedge R]), \text{ by Lemma 2.1,} \\
&\subseteq \mathrm{Hom}_r(\mathrm{Sub}[(\mathcal{L}' \wedge R), \mathcal{L} \wedge R]), \text{ since } \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}) \subseteq \mathcal{L}, \\
&\subseteq \mathrm{Sub}[\mathrm{Hom}_r(\mathcal{L}' \wedge R \wedge R), \mathrm{Hom}_r(\mathrm{Sub}(\mathcal{F}_o, \mathcal{L} \wedge R))], \text{ by Lemma 2.2,} \\
&\subseteq \mathrm{Sub}[\mathcal{L}_\infty, \mathrm{Hom}_r(\mathrm{Sub}(\mathcal{F}_o, \mathcal{L} \wedge R))], \\
&\quad \text{since } R \wedge R = R \text{ and } \mathrm{Hom}_r(\mathcal{L}' \wedge R) \subseteq \mathcal{L}_\infty.
\end{aligned}$$

Now

$$\begin{aligned}
\mathrm{Hom}_r[\mathrm{Sub}(\mathcal{F}_o, \mathcal{L} \wedge R)] &= \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{F}_o \wedge R, \mathcal{L} \wedge R)], \text{ since } \mathcal{F}_o \wedge R = \mathcal{F}_o, \\
&\subseteq \mathrm{Hom}_r \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{F}_o, \mathrm{Sub}(\mathcal{F}_o, \mathcal{L})) \wedge R], \text{ by Lemma 3.1,} \\
&\subseteq \mathrm{Hom}_r[\mathrm{Sub}(\mathcal{F}_o, \mathcal{L}) \wedge R], \text{ since } \mathrm{Sub}(\mathcal{F}_o, \mathcal{L}) \subseteq \mathcal{L} \\
&\qquad \text{and } \mathrm{Hom}_r \mathrm{Hom}_r = \mathrm{Hom}_r, \\
&\subseteq \mathrm{Hom}_r(\mathcal{L} \wedge R) \\
&\subseteq \mathcal{L}_\infty.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{Hom}_r[\mathrm{Sub}_n(\mathcal{L},\mathcal{L}) \wedge R] &\subseteq \mathrm{Sub}(\mathcal{L}_\infty, \mathcal{L}_\infty) \\
&\subseteq \mathcal{L}_\infty,
\end{aligned}$$

so that the induction is extended.

To complete the proof, we have

$$\begin{aligned}\mathrm{Hom}_r(\mathcal{L}_\infty \wedge \mathcal{R}) &= \mathrm{Hom}_r([\bigcup_{j=0}^{\infty}\mathrm{Sub}_j(\mathcal{L},\mathcal{L}) \wedge \mathcal{R}], \text{ by Theorem 4.1,}\\ &= \mathrm{Hom}_r(\bigcup_{j=0}^{\infty}[\mathrm{Sub}_j(\mathcal{L},\mathcal{L}) \wedge \mathcal{R}])\\ &\subseteq \mathcal{L}_\infty.\end{aligned}$$

(b) It suffices to show that "only if." Thus assume either $\mathrm{Hom}(\mathcal{L} \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$ or $\mathcal{L} \wedge \mathcal{R} \subseteq \mathcal{L}_\infty$. Clearly $\mathcal{R} \subseteq \mathcal{L}_\infty$. Thus $\mathcal{L}_\infty$ is closed under arbitrary homomorphism since it is closed under substitution. Hence $\mathrm{Hom}(\mathcal{L} \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$ if $\mathcal{L} \wedge \mathcal{R} \subseteq \mathcal{L}_\infty$. (Obviously, $\mathcal{L} \wedge \mathcal{R} \subseteq \mathcal{L}_\infty$ if $\mathrm{Hom}(\mathcal{L} \wedge \mathcal{R}) \subseteq \mathcal{L}_\infty$.) By (a), $\mathcal{L}_\infty$ is an AFL. Since $\mathcal{L}_\infty$ is closed under arbitrary homomorphism, it is a full AFL.

The above theorem gives another proof of the following result [6].

<u>Corollary</u>. The family of derivation-bounded languages is a full AFL.

<u>Proof</u>. It is known [6] that the family of derivation-bounded languages is the substitution closure of $\mathcal{L}_{lin}$, the family of linear context-free languages. Since $\mathcal{L}_{lin}$ contains $\mathcal{R}$ and is closed under intersection with regular sets, and $\mathrm{Sub}(\mathcal{F}_o, \mathcal{L}_{lin}) \subseteq \mathcal{L}_{lin}$, the substitution closure of $\mathcal{L}_{lin}$ is a full AFL by Theorem 4.4.

BIBLIOGRAPHY

[1] Bar-Hillel, Y., Perles, M., and Shamir, E. "On formal properties of simple phrase structure grammars," Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung, Vol. 14, 1961, pp. 143-172.

[2] Copi, I.M., Elgot, C.C., and Wright, J.B. "Realization of events by logical nets," Journal of the Association for Computing Machinery, Vol. 5, 1958, pp. 181-186.

[3] Ginsburg, S. and Greibach, S. "Abstract families of languages," to appear in a memoir of the American Mathematical Society.

[4] Ginsburg, S., Greibach, S., and Hopcroft, J. "Pre-AFL," to appear in a memoir of the American Mathematical Society.

[5] Ginsburg, S. and Rose, G.F. "Operations which preserve definability in languages," Journal of the Association for Computing Machinery, Vol. 10, 1963, pp. 175-195.

[6] Ginsburg, S. and Spanier, E.H. "Derivation-bounded languages," to appear in the Journal of Computer and System Sciences.

[7] Greibach, S. and Hopcroft, J. "Independence of AFL Operations," to appear in a memoir of the American Mathematical Society.

[8] Rabin, M.O. and Scott, D. "Finite automata and their decision problems," IBM Journal of Research and Development, Vol. 3, 1959, pp. 114-125.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) System Development Corporation 2500 Colorado Avenue Santa Monica, California	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

SUBSTITUTION IN FAMILIES OF LANGUAGES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Seymour Ginsburg
Edwin H. Spanier

6. REPORT DATE 16 September 1968	7a. TOTAL NO. OF PAGES 37	7b. NO. OF REFS 8
8a. CONTRACT OR GRANT NO. F1962867C0008 Grant No. AF-AFOSR-1203-67 Project, Task, Work Unit Nos. 5632-05-01 c DoD Element 61445O1F d DoD Subelement 681305	9a. ORIGINATOR'S REPORT NUMBER(S) TM-738/049/00 Scientific Report No. 22 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFCRL-68-0592	

10. DISTRIBUTION STATEMENT

1- Distribution of this document is unlimited. It may be released to the clearinghouse, Department of Commerce, for sale to the general public.

11. SUPPLEMENTARY NOTES Hq. AFOSR, OAR (SRIR) United States Air Force Arlington, Virginia	12. SPONSORING MILITARY ACTIVITY Air Force Cambridge Research Laboratories (CRB) L.G. Hanscom Field Bedford, Massachusetts 01730

13. ABSTRACT

The effect of substitution in families of languages, especially AFL, is considered. Among the main results shown are the following: The Substitution of one AFL into another is an AFL. Under suitable hypotheses, the AFL generated by the family obtained from the substitution of one family into another, is the family obtained from the substitution of the corresponding AFL. A condition is given for the AFL generated by the substitution closure of a family to be the substitution closure of the AFL generated by the family.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Families of languages						
Abstract family of languages (AFL)						
AFL						
Substitution closure						

Unclassified

Security Classification